UNIVERSITÄT BONN

Deep Learning for Predicting Impacts of Extreme Events and their Drivers

DISSERTATION

zur Erlangung des Doktorgrades (*Dr. rer. nat.*)

der Mathematisch-Naturwissenschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

MOHAMAD HAKAM SHAMS EDDIN

aus

Altal, Damaskus, Syrien

Bonn, 2025

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachter / Betreuer: Prof. Dr. Juergen Gall
Gutachter: Prof. Dr. Joachim Denzler

Tag der mündlichen Prüfung / Date of oral examination: 25.03.2026
Erscheinungsjahr / Year of Publication: 2026

Abstract

by Mohamad Hakam Shams Eddin

for the degree of

Doctor rerum naturalium

Climate change will apparently alter the patterns and frequencies of extreme weather and climate events. Yet, we still struggle to accurately localize and predict where extreme events and their impacts could occur in the future. If we were able to anticipate these events sooner, we could devise more informed decisions and far better adaptation strategies. Recognizing this, scientific efforts are dedicated to identify and forecast extreme events and their impacts using climate data. This dissertation explores applications of artificial neural networks for early warning and forecasting of extreme events and their impacts, such as wildfire, extreme floods and agricultural droughts. We address four main topics.

First, we introduce a novel deep learning approach for the next day wildfire danger forecasting. Previous approaches neglect the intrinsic difference between static and dynamic input variables. In our approach, we propose a 2D/3D two-branch Convolutional Neural Network (CNN) with Location-aware Adaptive Normalization (LOAN) to address this issue. The multi-branch architecture can handle multi-source data and the LOAN layer can modulate dynamic features using the corresponding static features as conditions. Thus, our approach for wildfire forecasting uses a unified 2D/3D model to handle distinct properties of input features.

Second, we address long-term agricultural drought forecasting and introduce Focal-TSMP, a deep learning approach that predicts satellite-derived vegetation indices from a regional climate simulation. We first use the climate simulation for long-term forecasting. Then, the proposed deep learning model translates these forecasts into Normalized Difference Vegetation Index (NDVI) and Brightness Temperature (BT) images that are not produced directly by the simulation. From these predicted images, we then derive key vegetation indices like Vegetation Health Index (VHI) as a proxy for agricultural drought. The proposed methodology is valuable for several downstream applications. Specifically, it allows us to fill gaps in satellite data by estimating NDVI and BT and provides a way to model how vegetation responds to extreme events across different climate change scenarios.

Third, there is a need of deep learning approaches to identify spatio-temporal relations between impacts of extremes and their drivers in high dimensional climate data. There is also a need of standardized benchmarks to tackle this challenging task. To address these issues, we introduce benchmarks and a novel deep learning approach to identify drivers of extreme agricultural droughts from reanalysis data. Our proposed approach is trained end-to-end to simultaneously predict spatio-temporally extreme impacts and their drivers in the input variables. We use a quantization technique to force the network to predict impacts conditioned on the spatio-temporal binary masks of identified drivers. This allows the network to successfully identify drivers that are correlated with the impacts of extremes.

Finally, we introduce RiverMamba for global river discharge and flood forecasting.

RiverMamba is a novel deep learning model that was developed to address the limitations of lumped and catchment concept approaches in hydrology. In our proposed approach, we leverage the spatio-temporal connections of bodies of water. RiverMamba achieves this by utilizing space-filling curves and efficient deep learning blocks to capture large-scale river networks. We build RiverMamba as a vision model that can be pretrained with reanalysis data and finetuned on observations to forecast floods on a 0.05° grid and up to 7 days lead time globally.

The proposed approaches outperform state-of-the-art and baseline methods in terms of effectiveness and efficiency, which we demonstrate through comprehensive evaluations on diverse benchmarks and datasets.

Keywords: extreme events, early warning systems, remote sensing, anomaly detection, artificial neural networks, Earth and climate sciences, flood forecasting, drought forecasting, wildfire forecasting

Acknowledgements

This dissertation took a long time to complete, and I am deeply grateful for the help and encouragement I received along the way.

First, a huge thank you to my advisor Prof. Dr. Juergen Gall for his guidance. Prof. Juergen Gall sharpened my critical thinking, reined in my digressions, and granted me the freedom to explore my research interests. I would also like to thank Prof. Dr. Joachim Denzler for dedicating the time to review this thesis.

Along the way a number of scientists have helped me shape my understanding of Earth science and corrected my misunderstandings. I would like to thank Dr. Petra Friederichs and Till Fohrmann for the thoughtful discussions on detection and attribution of weather and climate extremes.

I also owe a big thanks to my co-authors Prof. Dr. Ribana Roscher and Prof. Dr. Stefan Kollet, for the invaluable contributions and discussions. Thanks especially to my friend and colleague Yikui Zhang, who accompanied me on the journey of developing a flood forecasting system.

Thanks to all my friends and colleagues in the Computer Vision Group at the University of Bonn for the welcoming environment and insightful discussions: Yaser Souri, Andreas Doering, Olga Zatsarynna, Emad Bahrami, Julian Tanke, Vadim Sushko, Jinhui Yi, Federico Spurio, Serdar Ozsoy, Hamid Suleman, Syed Talal Wasim, Fatemeh Jabbari, Anas Al-lahham, Lars Doorenbos, Sina Mokhtarzadeh, Yanan Luo, Enrico Pallotta, Sassan Mokhtar, Shi-Jie Li, and Shuai Li. Thanks also to my friends and colleagues in the Humanoid Robots Lab at the University of Bonn for the thoughtful discussions: Murad Dawood and Ahmad Shokry. Thanks also to my friend Mohamad Yaghmour from the LIMES institute.

Finally, I would like to thank Dr. Helmut Rabenau and the Mayer Family: Andrea, Michael, Simon, Thomas, Michelle, and Tabea, who provided kind encouragement and a sense of home in Germany. To my family, my sister Darin and brothers Nebal, Sameh, and Naser, thanks for always being their for me. Thanks to my fiancée, Rokaya, for her consistent emotional support and patience. And most importantly, I am deeply indebted to my mother Nada for her unconditional love and care. And finally, to my father Dr. Naem Shams Eddin, who won't be here to see me graduate, thank you.

Feel by turns the bitter change
Of fierce extremes, extremes by change more fierce.

—JOHN MILTON, *Paradise Lost (1667)*

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Challenges . . . 3
  - 1.2.1 Multi-source Data . . . 3
  - 1.2.2 Imbalanced Data . . . 3
  - 1.2.3 Efficiency . . . 4
  - 1.2.4 Extreme Events Characteristics . . . 4
  - 1.2.5 Understanding of Extreme Events Impacts . . . 4
  - 1.2.6 Trustworthiness . . . 5
- 1.3 Contributions . . . 5
  - 1.3.1 Location-aware Adaptive Normalization: A Deep Learning Approach For Wildfire Danger Forecasting . . . 5
  - 1.3.2 Focal-TSMP: Deep Learning for Vegetation Health Prediction and Agricultural Drought Assessment from a Regional Climate Simulation . . . 6
  - 1.3.3 Identifying Spatio-Temporal Drivers of Extreme Events . . . 6
  - 1.3.4 RiverMamba: A State Space Model for Global River Discharge and Flood Forecasting . . . 8
- 1.4 Dissertation Structure . . . 9

2 Background **11**
- 2.1 Artificial Neural Network (ANN) . . . 12
  - 2.1.1 Multi Layer Perceptron (MLP) . . . 12
  - 2.1.2 Convolutional Neural Network (CNN) . . . 13
  - 2.1.3 Long Short-Term Memory (LSTM) . . . 14
  - 2.1.4 Convolutional Long Short-Term Memory (ConvLSTM) . . . 15
  - 2.1.5 Vision Transformer (ViT) . . . 15
  - 2.1.6 Focal Modulation Network . . . 17
  - 2.1.7 State Space Model (SSM) and the Mamba Family . . . 18
- 2.2 Datasets . . . 19
  - 2.2.1 FireCube . . . 19
  - 2.2.2 Terrestrial System Modelling Platform (TSMP) . . . 21
  - 2.2.3 Observational Remote Sensing Data . . . 23
  - 2.2.4 ERA5-Land Reanalysis Data . . . 27
  - 2.2.5 CERRA Reanalysis Data . . . 28
  - 2.2.6 GloFAS . . . 30
  - 2.2.7 GRDC Observational River Discharge Data . . . 32
  - 2.2.8 HRES Meteorological Forecast . . . 33
  - 2.2.9 CPC Precipitation Data . . . 34
  - 2.2.10 LISFLOOD Static Data . . . 35
  - 2.2.11 HydroRIVERS Static Data . . . 35
  - 2.2.12 Diagnostic River Points for Flood Forecasting . . . 36

2.3 Return Periods and Flood Definition . . . 36
2.4 PRUDENCE Scientific Regions . . . 38
2.5 Evaluation Metrics . . . 39
2.5.1 Mean Absolute Error (MAE) . . . 39
2.5.2 Root Mean Square Error (RMSE) . . . 39
2.5.3 Bias . . . 39
2.5.4 Pearson Correlation Coefficient (R_p) . . . 39
2.5.5 Spearman Correlation Coefficient (R_s) . . . 40
2.5.6 Coefficient of Determination (R2) . . . 40
2.5.7 Kling-Gupta Efficiency (KGE) . . . 40
2.5.8 Precision . . . 40
2.5.9 Recall . . . 41
2.5.10 F1-score . . . 41
2.5.11 Intersection over Union (IoU) . . . 41
2.5.12 Overall Accuracy (OA) . . . 41
2.5.13 Area Under the Receiver Operating Characteristic Curve (AUROC) . . . 41

3 Related Works 43
3.1 Anomaly Detection . . . 43
3.1.1 Anomalies and Extremes Detection in Climate Data . . . 43
3.1.2 Anomaly Detection Algorithms . . . 44
3.2 Extreme Events Forecasting . . . 45
3.2.1 Wildfire Danger Forecasting . . . 45
3.2.2 Vegetation Health and Agricultural Drought Forecasting . . . 46
3.2.3 River Discharge and Flood Forecasting . . . 47
3.3 Techniques in Neural Networks . . . 48
3.3.1 Conditional Normalization . . . 48
3.3.2 Multi-Branch Neural Networks . . . 49
3.3.3 Temporal Positional Encoding . . . 50
3.3.4 State Space Model (SSM) and Mamba . . . 50

4 Location-aware Adaptive Normalization: A Deep Learning Approach For Wildfire Danger Forecasting 51
4.1 Introduction . . . 52
4.2 Methodology . . . 53
4.2.1 2D/3D Two-Branch CNN . . . 54
4.2.2 Location-Aware Adaptive Normalization (LOAN) . . . 55
4.2.3 Absolute Temporal Encoding (TE) . . . 58
4.3 Experimental Results and Analysis . . . 58
4.3.1 Experimental Setup . . . 58
4.3.2 Comparison with Baselines . . . 60
4.3.3 Qualitative Results . . . 63
4.3.4 Variable Importance . . . 71
4.3.5 Comparing Different Conditional Maps . . . 71

4.3.6 Ablation Studies . . . 72
4.4 Summary . . . 75

5 Focal-TSMP: Deep Learning for Vegetation Health Prediction and Agricultural Drought Assessment from a Regional Climate Simulation 77
5.1 Introduction . . . 79
5.2 Methodology . . . 81
5.2.1 Model Architectures . . . 81
5.2.2 Focal Modulations . . . 83
5.2.3 Channel Attention . . . 86
5.2.4 Loss Function . . . 87
5.3 Experimental results and analysis . . . 87
5.3.1 Experimental Setup . . . 88
5.3.2 NDVI and BT Prediction . . . 89
5.3.3 Agricultural Drought Assessment . . . 98
5.3.4 Variable Importance . . . 101
5.3.5 Ablation Studies . . . 104
5.4 Summary . . . 106

6 Identifying Spatio-Temporal Drivers of Extreme Events 107
6.1 Introduction . . . 108
6.2 Method . . . 109
6.3 Synthetic data . . . 112
6.4 Experimental results and analysis . . . 115
6.4.1 Experimental Setup . . . 115
6.4.2 Generating Synthetic Data . . . 119
6.4.3 Experiments on Synthetic Data . . . 129
6.4.4 Ablation Studies . . . 133
6.4.5 Experiments on Real-world Data . . . 138
6.5 Summary . . . 151

7 RiverMamba: A State Space Model for Global River Discharge and Flood Forecasting 153
7.1 Introduction . . . 154
7.2 Method . . . 156
7.2.1 Hindcast Layer . . . 157
7.2.2 Hindcast Block . . . 158
7.2.3 Serialization . . . 159
7.2.4 Location-aware Adaptive Normalization Layers (LOAN) . . . 160
7.2.5 Mamba Block . . . 160
7.2.6 Forecasting Layer . . . 163
7.2.7 Training . . . 165
7.3 Experimental results . . . 166
7.3.1 Experimental Setup . . . 166
7.3.2 Experiments on Gauged GloFAS River Discharge Reanalysis . . . 171
7.3.3 Experiments on Gauged GRDC Observational River Discharge . . . 181

7.3.4 Comparison with Google Reforecast on Ungauged GRDC 192
7.3.5 Comparison with Google Reforecast on Gauged GRDC 195
7.3.6 Comparison to Operational GloFAS Forecsating on Gauged GRDC 195
7.3.7 Case Studies of Extreme Flood Events 198
7.3.8 Ablation Studies 203
7.4 Summary 209

8 Conclusion **211**
8.1 Summary 211
8.2 Limitations 213
8.2.1 Wildfire Forecasting 213
8.2.2 Agricultural Drought Forecasting 214
8.2.3 Identifying Drivers of Extreme Events Impacts 214
8.2.4 Flood Forecasting 214
8.3 Outlook and Future Works 215
8.3.1 Ensemble Forecast 215
8.3.2 Forecasting Vegetation Conditions from Climate Scenarios 216
8.3.3 Agricultural Drought Forecasting 217
8.3.4 Identifying Drivers of Extreme Events Impacts 217
8.3.5 Flood Forecasting 218

List of Figures

1.1 An example of wildfire forecasting with deep learning. 6
1.2 An overview of the proposed task with the Focal-TSMP model. 7
1.3 An overview of the task of identifying spatio-temporal drivers of extremes impacts. . 7
1.4 An overview of the task of medium-range river discharge and flood forecasting with deep learning. 8

2.1 Illustrative comparison between window-based self-attention and focal modulation. . 17
2.2 Comparison between self-attention and focal modulation 18
2.3 An overview of the extreme agricultural droughts definition from remote sensing. . . 24
2.4 The definition of the domains used in the dissertation. 28
2.5 An overview of the key modules in the GloFAS forecasting system. 31
2.6 Locations of the selected 3366 GRDC stations. 33
2.7 Hydrograph for GloFAS-reanalysis and observations from 1 January 2000 to 31 December 2023. 34
2.8 An overview of the processed HydroRIVERS static features. 36
2.9 Overview of the selected river points. 37
2.10 Orography over the EURO-CORDEX domain. 38

4.1 Overview of the proposed approach for wildfire forecasting. 54
4.2 Illustration of the Location-aware Adaptive Normalization layer (LOAN) when using the conditional map from the static branch. 56
4.3 Illustration of the Location-aware Adaptive Normalization layer (LOAN) when using C static variables directly for modulation. 57
4.4 Illustration of how the absolute temporal encoding is added in the model. 59
4.5 Distribution of the wildfire events per day between $06/03/2009$ and $29/08/2021$ (FireCube dataset). 59
4.6 Predictions for 8 consecutive days in August. 64
4.7 Qualitative results for 6 days during the wildfire season in year 2020. 65
4.8 Qualitative results for 6 days during the wildfire season in year 2020. 66
4.9 Qualitative results for 6 days during the wildfire season in the extreme year 2021. . . 67
4.10 Qualitative results for 6 days during the wildfire season in the extreme year 2021. . . 68
4.11 Impact of the number of testing samples on the precision. 74
4.12 BCE loss during training for the training and validation set. 74

5.1 An overview of the proposed model to predict NDVI and BT from a TSMP climate simulation. 82
5.2 An illustration of the focal modulation block. 83
5.3 An illustration of the function FM. 85
5.4 Illustration of the channel attention implementation. 86
5.5 Examples predictions for weekly NDVI from the test set. 92
5.6 Examples predictions for weekly BT from the test set. 93

5.7 An analysis of uncertainty and model generalization for different times of the year. . 94
5.8 An analysis of uncertainty and model generalization for different time of the year over each PRUDENCE region. 95
5.9 Comparison of spatially averaged weekly agricultural drought indices between the model prediction and NOAA observation over each PRUDENCE region. 96
5.10 Comparison of spatially averaged weekly agricultural drought indices between the model prediction and NOAA observations over each PRUDENCE region. 97
5.11 Supplementary results to Fig. 5.9. 98
5.12 Supplementary results to Fig. 5.10. 99
5.13 Comparison between the seasonal predicted Vegetation Health Index (VHI) and NOAA observations over Pan-Europe domain. 100
5.14 Comparison between the predicted drought frequency and NOAA observations over Pan-Europe domain. 100
5.15 An evaluation of seasonally predicted agricultural drought indices with ground truth NOAA observations at the resolution 0.88°. 101
5.16 Channel attention for TSMP input variables. 103
5.17 Impact of the perceptual VGG loss on NDVI predictions and image sharpness. 105
5.18 Impact of the perceptual VGG loss on BT predictions and image sharpness. 105

6.1 Overview of identifying spatio-temporal relations between the impacts of extremes and their drivers. 109
6.2 An overview of the proposed model to identify the spatio-temporal relations between extreme agricultural droughts and their drivers. 110
6.3 An example of the randomly generated coupling matrix between the synthetic variables and extremes for the synthetic CERRA dataset. 113
6.4 Examples of the synthetic CERRA reanalysis data described in Table 6.1. 120
6.5 Examples of the synthetic NOAA remote sensing described in Table 6.2. 121
6.6 Examples of the synthetic artificial data described in Table 6.3. 122
6.7 Visualization of the generated signals $\mathbf{\Phi}$ for 6 different variables from the synthetic CERRA reanalysis. 125
6.8 Visualization of the generated signals $\mathbf{\Phi}$ for 5 different variables from the synthetic NOAA data. 126
6.9 Visualization of the generated signals $\mathbf{\Phi}$ for 6 different variables from the synthetic artificial data. 127
6.10 Qualitative results on the synthetic CERRA reanalysis from the test set. 132
6.11 F1-score with different correlation settings between the input variables and extreme events. 132
6.12 Supplementary to the ablation study in Table 6.12. 136
6.13 Temporal evolution of drivers during the extremes. 140
6.14 Qualitative results on ERA5-Land over the EUR-11 domain. 142
6.15 The averaged spatial distribution of drivers and anomalies related to Portugal in Europe. 143
6.16 The averaged spatial distribution of drivers and anomalies related to a specific place in Europe (North Rhine-Westphalia). 143
6.17 Qualitative results on ERA5-Land for Europe (EUR-11). 144

6.18 Temporal evolution of drivers and anomalies related to the extremes in ERA5-Land for Europe (EUR-11). . . . 144
6.19 Qualitative results on ERA5-Land for Africa (AFR-11). . . . 145
6.20 Temporal evolution of drivers and anomalies related to the extremes in ERA5-Land for Africa (AFR-11). . . . 145
6.21 Qualitative results on ERA5-Land for South America (SAM-11). . . . 146
6.22 Temporal evolution of drivers and anomalies related to the extremes in ERA5-Land for South America (SAM-11). . . . 146
6.23 Qualitative results on ERA5-Land for North America (NAM-11). . . . 147
6.24 Temporal evolution of drivers and anomalies related to the extremes in ERA5-Land for North America (NAM-11). . . . 147
6.25 Qualitative results on ERA5-Land for East Asia (EAS-11). . . . 148
6.26 Temporal evolution of drivers and anomalies related to the extremes in ERA5-Land for East Asia (EAS-11). . . . 148
6.27 Qualitative results on ERA5-Land for Central Asia (CAS-11). . . . 149
6.28 Temporal evolution of drivers and anomalies related to the extremes in ERA5-Land for Central Asia (CAS-11). . . . 149
6.29 Qualitative results on CERRA reanalysis for Europe. . . . 150
6.30 Temporal evolution of drivers and anomalies related to the extremes in CERRA reanalysis for Europe. . . . 150

7.1 Example of a 5-day forecast of river discharge and flood events. . . . 156
7.2 An overview of the proposed RiverMamba model for river discharge forecasting. . . . 158
7.3 Illustration of the spatial scans in RiverMamba. . . . 159
7.4 Illustration of the difference between global curve and grouped local curves. . . . 160
7.5 Visualization of different types of space-filling curves. . . . 161
7.6 Visualization of different types of space-filling curves. . . . 162
7.7 The structure of the hindcast block and forecast block. . . . 163
7.8 A simplified view of splitting along Gilbert space-filling curve. . . . 166
7.9 Details about the data splits for RiverMamba. . . . 169
7.10 Results on GloFAS reanalysis across lead times and flood return periods. . . . 172
7.11 MAE of the river discharge forecasting with different lead time on GloFAS reanalysis. 173
7.12 RMSE of the river discharge forecasting with different lead time on GloFAS reanalysis. 173
7.13 R2 (NSE) of the river discharge forecasting with different lead time on GloFAS reanalysis. . . . 173
7.14 Pearson correlation (R) of the river discharge forecasting with different lead time on GloFAS reanalysis. . . . 173
7.15 F1-score of flood forecasting for different return periods and lead time on GloFAS reanalysis. . . . 174
7.16 Precision of flood forecasting for different return periods and lead time on GloFAS reanalysis. . . . 175
7.17 Recall of flood forecasting for different return periods and lead time on GloFAS reanalysis. . . . 176

7.18 F1-score of flood forecasting for different lead time and return periods on GloFAS reanalysis. . . . 177
7.19 Precision of flood forecasting for different lead time and return periods on GloFAS reanalysis. . . . 178
7.20 Recall of flood forecasting for different lead time and return periods on GloFAS reanalysis. . . . 179
7.21 Comparison of F1-score between RiverMamba and LSTM on GloFAS reanalysis for the 5-year return period events. . . . 180
7.22 Comparison of KGE between RiverMamba and LSTM on GloFAS reanalysis. . . . 180
7.23 Comparison to LSTM on GloFAS reanalysis. . . . 181
7.24 Results on gauged GRDC observations across lead times and flood return periods. . . . 182
7.25 MAE of the river discharge forecasting with different lead time on GRDC observations. 183
7.26 RMSE of the river discharge forecasting with different lead time on GRDC observations. . . . 183
7.27 R2 (NSE) of the river discharge forecasting with different lead time on GRDC observations. . . . 183
7.28 Pearson correlation (R) of the river discharge forecasting with different lead time on GRDC observations. . . . 183
7.29 Comparison to LSTM on GRDC observations. . . . 184
7.30 Comparison to GloFAS (ECMWF) - reforecast on GRDC observations. . . . 184
7.31 F1-score of flood forecasting for different return periods and lead time on GRDC observations. . . . 185
7.32 Precision of flood forecasting for different return periods and lead time on GRDC observations. . . . 186
7.33 Recall of flood forecasting for different return periods and lead time on GRDC observations. . . . 187
7.34 F1-score of flood forecasting for different lead time and return periods on GRDC observations. . . . 188
7.35 Precision of flood forecasting for different lead time and return periods on GRDC observations. . . . 189
7.36 Recall of flood forecasting for different lead time and return periods on GRDC observations. . . . 190
7.37 Comparison of F1-score between RiverMamba and LSTM (Google) on GRDC observations for the 2-year return period events. . . . 191
7.38 Comparison of F1-score between RiverMamba and GloFAS reforecast on GRDC observations for the 2-year return period events. . . . 191
7.39 Comparison of KGE between RiverMamba and LSTM (Google) on GloFAS reanalysis. 192
7.40 Comparison to Google reforecast. Shown is F1-score of flood forecasting for different return periods and lead time on ungauged GRDC stations. . . . 193
7.41 Comparison to Google reforecast on ungauged GRDC stations. . . . 194
7.42 Comparison to Google reforecast. Shown is F1-score of flood forecasting for different lead time and return periods on ungauged GRDC stations. . . . 194
7.43 Comparison to Google reforecast on gauged GRDC stations. . . . 195

7.44 Comparison to Google reforecast. Shown is F1-score of flood forecasting for different return periods and lead time on gauged GRDC stations. 196
7.45 Comparison to Google reforecast. Shown is F1-score of flood forecasting for different lead time and return periods. 197
7.46 River discharge of the Sauer river in 2021. 198
7.47 Daily total precipitation from 2021-07-11 to 2021-07-17 during the 2021 flood event in the target domain. 199
7.48 Comparison between GloFAS reanalysis and RiverMamba forecast during the Southeast European flood in October 2024. 199
7.49 Comparison between GloFAS reanalysis and RiverMamba forecast during the Central European floods event in 2024. 200
7.50 Comparison between GloFAS reanalysis and RiverMamba forecast during the Spanish flood event in October 2024. 200
7.51 Comparison between GloFAS reanalysis and RiverMamba forecast during the Saarland Germany flood event in May 2024. 201
7.52 Comparison between GloFAS reanalysis and RiverMamba forecast during the Kenya-Tanzania flood event in April 2024. 201
7.53 Comparison between GloFAS reanalysis and RiverMamba forecast during the California, USA flood in February 2024. 202
7.54 Comparison between GloFAS reanalysis and RiverMamba forecast during the Central-South China flood event in June 2024. 202
7.55 Comparison between Mamba and Transformer-based backbones. 205
7.56 Inference time for RiverMamba with respect to the length of input tokens. 208

8.1 An overview of the probabilistic generative process of RiverMamba 216
8.2 An illustration of testing the developed model for land- and water use changes scenarios. 217
8.3 An overview of the same product output generated by different hydrological systems. 219

List of Tables

2.1 The static variables used in FireCube dataset. . . . 20
2.2 The dynamic variables used in FireCube dataset. . . . 20
2.3 Technical details on the output variables in the TSMP EUR-11 simulation. . . . 22
2.4 Technical details on the static variables from CLM in the TSMP EUR-11 simulation and the computed static variables. . . . 23
2.5 Technical details on the spectral channel characteristics for Advanced Very High Resolution Radiometer (AVHRR) and Visible Infrared Imaging Radiometer Suite (VIIRS). . . . 24
2.6 Details regarding the ratio of extreme events in the pre-processed NOAA remote sensing data. . . . 27
2.7 Datasets used in the experiments on real-world data. . . . 28
2.8 Details about the processed variables from ERA5-Land reanalysis in Chapter 7. . . . 29
2.9 Details regarding the processed variables from ERA5-Land reanalysis in Chapter 6. . 30
2.10 CERRA reanalysis used in the experiments on real-world data in Chapter 6. . . . 30
2.11 Details regarding the processed variables from CERRA reanalysis used in Chapter 6. 31
2.12 Details about the processed variables from GloFAS reanalysis. . . . 32
2.13 Details about the processed variables from the ECMWF Integrated Forecast System (IFS) High Resolution (HRES) atmospheric model. . . . 34
2.14 The processed LISFLOOD static and parameter maps. . . . 35

4.1 Comparison with baselines on the FireCube dataset. . . . 61
4.2 Quantitative results of different deep learning model designs on the FireCube dataset. 62
4.3 Comparison between different deep learning model designs on the FireCube dataset. 63
4.4 Recall (%) metric per class for Figs. 4.7-4.10. . . . 69
4.5 Performance metrics for different categories of static variables. . . . 70
4.6 Impact of different conditional maps on the feature modulation. . . . 72
4.7 Ablation study of different position choices for the LOAN layer. . . . 73

5.1 Comparing the performance of different DL models for NDVI and BT prediction on the validation set. . . . 90
5.2 Comparing the performance of different DL models for NDVI and BT prediction on the test set. . . . 91
5.3 Inference time in seconds for different DL models. . . . 91
5.4 Impact of TSMP model components on the model performance. . . . 102
5.5 Ablation study on the perceptual VGG loss described in Eq. (5.11). . . . 104
5.6 Impact of patch size for patch embedding before the first encoder layer. . . . 106

6.1 Configurations of the synthetic CERRA reanalysis. . . . 123
6.2 Configurations of the synthetic NOAA remote sensing. . . . 123
6.3 Configurations of the synthetic artificial data. . . . 124
6.4 Driver detection results on the synthetic CERRA reanalysis. . . . 128

6.5 Driver detection results on the synthetic NOAA reanalysis. 130
6.6 Driver detection results on the synthetic artificial data. 131
6.7 Inference time in seconds for our model and other DL baselines. 133
6.8 Training time on the synthetic CERRA for 1 epoch. 134
6.9 Ablation studies on the key model architecture from the validation set. 134
6.10 Ablation study on the backbone f_θ used for feature extraction. 135
6.11 Ablation studies on the temporal resolution from the validation set. 135
6.12 Ablation studies on the loss function from the validation set. 135
6.13 Ablation studies on the loss function $\mathcal{L}_{(quantize)}$ in Eq. (6.5). 136
6.14 Ablation studies on the weighted loss function $\mathcal{L}_{(extreme)}$ in Eq. (6.4). 137
6.15 Sensitivity studies on the weighting parameters used in the main loss function in Eq. (6.7). 137
6.16 Ablation studies on the quantization layer. 138
6.17 Quantitative results for extreme drought detection on real-world data. 139
6.18 Ablation study on the driver detection step. 139
6.19 Quantitative results from ERA5-Land EUR-11 data for different combination of physical variables. 141
6.20 The relation between the definition of extremes from VHI and the model prediction. . 141
6.21 The impact of spatial resolution on the model prediction. 142

7.1 Implementation details of RiverMamba 168
7.2 Implementation details of the LSTM model. 170
7.3 Results on GloFAS-Reanalysis. 172
7.4 Results on GRDC gauged stations. 181
7.5 Comparison to Google reforecast on ungauged GRDC stations for the years 2014-2021. 193
7.6 Comparison to Google reforecast on gauged GRDC stations for the years 2019-2021. 195
7.7 Comparison to Operational GloFAS Forecsat on Gauged GRDC. 195
7.8 Ablation study on the loss function on the validation set over Europe. 203
7.9 Ablation study on the validation set over Europe regarding the weighting in the objective function. Shown is F1-score for reanalysis data across different return periods. 203
7.10 Ablation study on the location embedding on the validation set over Europe. 204
7.11 Ablation study on the forecasting strategy in RiverMamba on the validation set over Europe. 204
7.12 Training time on the reanalysis data for 1 epoch. 205
7.13 Ablation studies for the RiverMamba backbone on the validation set over Europe. . . 205
7.14 Ablation studies for variable importance on the validation set over Europe. 206
7.15 Ablation studies for pretraining on GloFAS reanalysis on the validation set over Europe. 206
7.16 Ablation studies for space-filling curves on the validation set over Europe. 206
7.17 Ablation regarding the spatio-temporal modeling on the validation set over Europe. . 207
7.18 Ablation studies for scan patterns on the validation set over Europe. 207
7.19 Ablation studies on the validation set over Europe. 207
7.20 Ablation study on the validation set over Europe regarding the activation function in LOAN layers. 208
7.21 Inference time in seconds. 208

Nomenclature

Abbreviations

An alphabetically sorted list of abbreviations used in the dissertation:

1D	One Dimensional
2D	Two Dimensional
3D	Three Dimensional
AFR	CORDEX Africa Domain
AI	Artificial Intelligence
AL	Alps Region
ANN	Artificial Neural Network
AUROC	Area Under the Receiver Operating Characteristic Curve
AVHRR	Advanced Very High-Resolution Radiometer
BI	British Isles Region
BT	Brightness Temperature
CAS	CORDEX Central Asia Domain
CDF	Cumulative Distribution Function
CLC	Copernicus Corine Land Cover
CLM	Community Land Model
CNN	Convolutional Neural Network
ConvLSTM	Convolutional Long Short-Term Memory
COSMO	Consortium for Small-scale Modeling
CPC	Climate Prediction Center
CPU	Central Processing Unit
DEM	Digital Elevation Model
DL	Deep Learning
DWConv	Depth-wise Convolution
EA	Eastern Europe Region
EAS	CODEX East Asia Domain
ECMWF	European Centre for Medium-Range Weather Forecasts
EFAS	European Flood Awareness System
EUR	CORDEX Europe Domain
FFL	Feed-Forward Layer
FM	Focal Modulation
FN	False Negative
FP	False Positive
FR	France Region
FWI	Fire Weather Index
GAP	Global Average Pooling
GELU	Gaussian Error Linear Unit
GloFAS	Global Flood Awareness System

GNN	Graph Neural Network
GPU	Graphics Processing Unit
GRDC	Global Runoff Data Centre
GRU	Gated Recurrent Units
HBM	High Bandwidth Memory
HRES	High Resolution Forecasts
IF	Isolated Forest
IFS	Integrated Forecast System
IG	Integrated Gradients
IMERG	Integrated Multi-satellite Retrievals for Global Precipitation Measurement
IoU	Intersection over Union
IP	Iberian Peninsula Region
KGE	Kling-Gupta Efficiency
LAI	Leaf Area Index
LFQ	Lockup Free Quantization
LOAN	Location-aware Adaptive Normalization
LSTM	Long Short-Term Memory
LULC	Land Use Land Cover
MAE	Mean Absolute Error
MD	Mediterranean Region
ME	Mid-Europe Region
MIL	Multiple Instance Learning
ML	Machine Learning
MLP	Multi Layer Perceptron
NAM	CORDEX North America Domain
NDVI	Normalized Difference Vegetation Index
NLP	Nature Language Processing
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
OA	Overall Accuracy
ParFlow	Parallel Flow
ReLU	Rectified Linear Unit
RF	Random Forests
RGB	Red-Green-Blue
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
S2S	Sub-Seasonal to Seasonal
SAM	CORDEX South America Domain
SC	Scandinavia Region
SRAM	Static Random-Access Memory
SSM	State Space Model
TCI	Thermal Condition Index
TE	Temporal Encoding
TN	True Negative

TP	True Positive
TSMP	Terrestrial Systems Modelling Platform
UTC	Universal Time Coordinated
VCI	Vegetation Condition Index
VGG	Visual Geometry Group
VHI	Vegetation Health Index
VIIRS	Visible Infrared Imaging Radiometer Suite
ViT	Vision Transformer
VP	Vegetation Product
VQ	Vector Quantization
WGS 84	World Geodetic System 1984

List of Publications

The dissertation is based on the following publications:

- **Location-aware Adaptive Normalization: A Deep Learning Approach For Wildfire Danger Forecasting**
  Mohamad Hakam Shams Eddin, Ribana Roscher, and Juergen Gall
  IEEE Transactions on Geoscience and Remote Sensing (TGRS), 2023
  Doi: 10.1109/TGRS.2023.3285401

- **Focal-TSMP: Deep Learning for Vegetation Health Prediction and Agricultural Drought Assessment from a Regional Climate Simulation**
  Mohamad Hakam Shams Eddin, and Juergen Gall
  Geoscientific Model Development (GMD), 2024
  Doi: 10.5194/gmd-17-2987-2024

- **Identifying Spatio-Temporal Drivers of Extreme Events**
  Mohamad Hakam Shams Eddin, and Juergen Gall
  Advances in Neural Information Processing Systems 38 (NeurIPS), 2024
  Doi: 10.48550/arXiv.2410.24075

- **RiverMamba: A State Space Model for Global River Discharge and Flood Forecasting**
  Mohamad Hakam Shams Eddin, Yikui Zhang, Stefan Kollet, and Juergen Gall
  Advances in Neural Information Processing Systems 39 (NeurIPS), 2025
  Doi: 10.48550/arXiv.2505.22535

CHAPTER 1

Introduction

Contents

1.1 Motivation **1**
1.2 Challenges **3**
1.2.1 Multi-source Data 3
1.2.2 Imbalanced Data 3
1.2.3 Efficiency 4
1.2.4 Extreme Events Characteristics 4
1.2.5 Understanding of Extreme Events Impacts 4
1.2.6 Trustworthiness 5
1.3 Contributions **5**
1.3.1 Location-aware Adaptive Normalization: A Deep Learning Approach For Wildfire Danger Forecasting 5
1.3.2 Focal-TSMP: Deep Learning for Vegetation Health Prediction and Agricultural Drought Assessment from a Regional Climate Simulation 6
1.3.3 Identifying Spatio-Temporal Drivers of Extreme Events 6
1.3.4 RiverMamba: A State Space Model for Global River Discharge and Flood Forecasting 8
1.4 Dissertation Structure **9**

1.1 Motivation

Patterns observed over the past and climate projections indicate that larger regions face a rising risk of severe weather and climate events, including heavy rainfall (*Prein et al.*, 2017), extreme floods (*Blöschl et al.*, 2017, 2019), extreme heatwaves (*Perkins-Kirkpatrick and Lewis*, 2020), and flash droughts (*Christian et al.*, 2023). It is generally anticipated that climate change will further reshape the future by altering how often, how long, and how intense extreme weather and climate events become (*Seneviratne et al.*, 2021). Since this has a significant impact on various sectors of our life, forecasting of extreme climate and weather events has become more necessary than ever.

Extremes forecasting refers to the task of providing spatio-temporal susceptibility maps that model a future state within the Earth system. A common approach is to forecast the extreme event itself, e.g., forecasting a wildfire event. Another common approach is to forecast a state variable as a proxy of extremes like temperature to model heatwaves or river discharge for flood forecasting. Typical methods to forecast such state variables rely on physics-based Numerical Weather Prediction (NWP) and climate models. These models are based on fundamental physical laws of motion and

energy. They bridge the gap between scientists, who are primarily concerned with understanding patterns of the past and present, and politicians and stakeholders, whose main interest lies in knowing the future. You might be surprised to know that patterns seen in the NWP and climate models are not programmed at all! Instead, they emerge when we accurately represent the underlying physics, and when the models successfully couple the Earth's subsystems simulating how energy and mass circulate the globe (*Easterbrook*, 2023). NWP can forecast short-term weather conditions, whereas climate models are built to mimic historical climate phenomena and they are also extended to explore future scenarios. To identify regions most vulnerable to extreme events, e.g., as a result of human activities under climate scenarios, we must answer the "what if" question.

However, weather on Earth is a chaotic system, and consequently, extreme forecasting is challenging by nature. If we want to forecast a particular state of a chaotic system, our forecast will be dependent on two main things: how precisely we know the initial conditions, and how good we can model the physical laws that govern the system and determine how each state leads to the next (*Lorenz*, 1993). Even with a flawless physics-based model, tiny errors in the initial state will limit how reliable we can predict the future (*Lorenz*, 1963). Getting the physics right is a modeling challenge, while improving the initial condition is more a compute-bound problem linked to observations. A model that takes less time to run, can start from a more accurate initial condition, which is the key to produce higher-quality forecasts (*Kalnay et al.*, 2024). Thus, having fast and efficient forecasting systems is more desirable than ever. Physics-based models have been developed rapidly over the past, whereas their computational demand is still a major challenge. For fast-onset events like storms and floods, resolving convection with NWP is still a computational bottleneck for an effective response (*Reichstein et al.*, 2025). Even with slow-onset events like droughts, Sub-seasonal to Seasonal (S2S) forecast is still highly uncertain (*Reichstein et al.*, 2025). Moreover, physics-based models often struggle to accurately predict extreme events such as extreme floods. This is because hydrological modeling still suffers from many unresolved processes, making conventional flood forecasting challenging (*Reichstein et al.*, 2025). To improve flood forecasting, we would need precise models that are capable of forecasting key hydrological variables like river discharge more accurately and much faster.

Deep learning models (*Reichstein et al.*, 2019) have showed that they are capable to learn complex relationships between predictors from climate data and the forecast variables, in an end-to-end manner (data-driven). Additionally, deep learning models can effectively handle multivariate and high-dimensional datasets like climate simulation and re-analysis data, which are typically used as initial conditions for forecasts. Furthermore, models based on deep learning can be used to handle biases implicitly, thus simplifying the entire workflow (*Schultz et al.*, 2021). Deep learning approaches have been therefore proposed for forecasting extreme weather and climate events, including extreme precipitation forecasting (*Gao et al.*, 2023b), tropical cyclone forecasting (*Lam et al.*, 2022), and flood forecasting (*Nearing et al.*, 2024).

While extreme events forecasting focuses on the extremes, impacts assessment can be also modeled with deep learning, which will be in the heart of this dissertation. The focus here is on the outcomes of preceding events (*Camps-Valls et al.*, 2025). To model impacts of extremes, usually an indicator is used to localize the impacts. For example, an indicator like poor vegetation health is used as a proxy of extreme agricultural droughts. In this dissertation, we introduce new techniques based on deep learning that forecast diverse extreme events and their impacts. In particular, we will address the wildfire forecasting, agricultural drought forecasting, and flood forecasting.

Yet our discussion above was about questions like "what will happen", "when will an event happen", and "where might an event happen". However, questions such as "why did an event happen" or "what had happened" are equally important. The latter help us understand how extreme events and their impacts unfold, and potentially identify the underlying factors that drive these events. In this dissertation, we thus also aim to learn the relations between the impacts of extreme events and their relevant drivers from a data-driven perspective. It is worth to note that spatio-temporal relations within the Earth system can also be inferred by causal inference and causal representation learning (*Runge et al.*, 2019a; *Schölkopf et al.*, 2021; *Brehmer et al.*, 2022; *Tibau et al.*, 2022; *Boussard et al.*, 2023; *Runge et al.*, 2023; *Yehezkel Rohekar et al.*, 2023; *Tárraga et al.*, 2024). In contrast to statistical methods (*Vogel et al.*, 2021), data-driven methods do not require a prior hypothesis about drivers for extremes. Instead, they generate hypotheses that can be verified by statistical methods in a second step. We believe that this is an important direction since climate reanalysis provides huge amount of data and it is infeasible to test all combinations. This is also known as a curse of dimensionality in causal discovery problems (*Runge et al.*, 2019b) and data-driven approaches are therefore needed to generate potential candidates.

The next two sections give an overview of the challenges (Section 1.2) and our contributions (Section 1.3).

1.2 Challenges

In this section, we will discuss some challenges that need to be considered when using deep learning for predicting extreme events and identifying their drivers.

1.2.1 Multi-source Data

Typically, predicting extreme events and identifying their impacts involve dealing with data from various sources. For instance, remote sensing observational data are commonly used in a combination with static geographic attributions like land use land cover (LULC) or topographic features. In addition, high dimensional reanalysis data like ERA5 (*Hersbach et al.*, 2020) or ERA5-Land (*Muñoz Sabater et al.*, 2021) have become a typical starting point for forecasting weather and extreme events. Modeling the terrestrial hydrologic and energy cycles requires also input variables from subsurface to the top of the atmosphere. Crucially, data that are being used as input to deep learning models have usually different mapping and temporal and spatial resolution. This is especially important to address when observational data are combined with output from simulations or reanalysis data. For example, the task of flood forecasting requires the combination of sparse gauged river observations with high-resolution land surface and weather forecast data. Thus, developing deep learning approaches for extreme events prediction should take into the consideration the recurring challenge of having data that come from multiple sources. Ideally, the developed deep learning model should be able to integrate and extract relevant features across diverse sources and scales of data.

1.2.2 Imbalanced Data

Extreme events and their impacts are by nature statistically undersampled and rare. Given this rarity, a major challenge when predicting extremes is the absence of sufficient data samples for training

and evaluating deep learning approaches. This is a well-known challenge of class imbalance in deep learning, which makes the predictive model prone to overfitting toward the most dominant class. In other words, the model will perform well for the overrepresentated classes, while its prediction quality of rare classes remains poor. This is also evidenced with regression tasks, where extreme values at the tails of the distribution become typically underrepresented by the model predictions. Moreover, the spatial distribution of the data, especially, for observational data is rather inhomogeneous. Think for example of the task of flood forecasting. The uneven global distribution of gauged river stations poses challenges for model training in data-scarce regions. Overcoming these challenges is an essential step to build a skillful model.

1.2.3 Efficiency

One of the primary applications of deep learning is to integrate and analyze huge amounts of climate and weather data. A major bottleneck here is the compute and memory requirement to conduct the analysis. To give an idea about typical properties of climate data, consider for example an image generation task. For a common dataset like ERA5-Land reanalysis, the resolution of an image at one time step is $\sim 3600 \times 1800$ with tens of variables compared to typical RGB image with 3 channels. In addition, the temporal coverage of reanalysis data can go up to decades in the past. Unfortunately, localizing extremes becomes more accurate and sometimes only possible with high spatio-temporal resolutions. Thus, for real-world applications of deep learning, it is important to design compute- and memory-efficient models. In addition, as mentioned in Section 1.1, better forecasts depend on the quality of the initial condition. Therefore, having efficient forecasting systems with fast inference would give scientists more time to run diagnostic analysis steps to improve the initial condition for the forecasting, consequently, improving the accuracy of forecasts.

1.2.4 Extreme Events Characteristics

Depending on the characteristics of the extreme event and the state variable, the forecasts can span a wide range of lead times, including short-range, medium-range, and sub-seasonal to seasonal (S2S) forecasting. Modeling challenges differ with respect to the lead time. A well-known challenge for medium-range and seasonal forecasting is the error accumulation in deep learning models over long time horizons. This leads usually to blurred and unrealistic forecasts of extreme events. Another modeling challenge is related to the scale, whether the forecast is on a small scale like regional modeling or large scale global forecast, which requires a generalized model across diverse locations. Consequently, developing deep learning models depends on what state variable the model aims to forecast, the scale, and the lead times. Currently, each forecasting task is handled by a specialized model.

1.2.5 Understanding of Extreme Events Impacts

In order to understand how impacts of extreme events develop, we have first to understand the drivers behind them. In general, the evolution of extreme events impacts like extreme agricultural droughts arises from complex interactions between land and atmospheric Earth state variables. The spatio-temporal relationships between extreme impacts and their drivers are characterized by seasonal variations in time lags and inhomogeneous spatial responses. A frontier challenge is therefore to identify

the spatio-temporal relations between extreme impacts and their drivers. Selecting appropriate indicators for a realistic and adequate modeling of extremes requires also an understanding of the relationships between drivers and extremes. However, such spatio-temporal relations in climate data are not fully understood. Another issue is that, while impacts of extreme events can be localized, e.g., via remote sensing measurements or reports in databases, there is a lack of sufficient expert annotations of their drivers in real-world data. This poses challenges for training and quantitatively evaluating deep learning models to identify driver in real-world data.

1.2.6 Trustworthiness

Extreme events prediction is a complex problem, which requires in most cases sophisticated models. Another challenge is related to the deployment of deep learning models. Stakeholders would rely on output of these models to operate early waring systems and to make decisions that ensure public safety. Thus, it is crucial to develop not just complex models, but also reliable and trustworthy ones. For instance, it is necessary to extensively evaluate the models performance and highlight where, when and why the models might fail, leading to inaccurate forecasts. Moreover, climate reanalysis provides vast amount of data that the models would use as predictors. In this regard, interpreting the model predictions, e.g., via analyzing the feature importance is important to build trust in decision-making and to advance our understanding of the underlying mechanisms behind extreme events.

1.3 Contributions

This dissertation proposes approaches to address the challenges discussed above. Our contributions, which are outlined below, address forecasting when the weather conditions would have a tendency to cause a wildlife event (Section 1.3.1), forecasting vegetation health and extreme agricultural droughts based on a climate simulation (Section 1.3.2), identifying anomalies and drivers of impacts on vegetation health caused by extreme weather and climate events (Section 1.3.3), and forecasting river discharge and floods globally and with high resolution (Section 1.3.4). By forecasting extreme wildfires, agricultural droughts and floods with higher accuracy, we hope that our contributions can play an effective role in adaptation planning and climate risk reduction. The long-term goal is to develop deep learning approaches that can be translated into solutions related to globally important issues like climate change and environmental risks. Below we summarize our main contributions.

1.3.1 Location-aware Adaptive Normalization: A Deep Learning Approach For Wildfire Danger Forecasting

In our first contribution, we introduce a 2D/3D Convolutional Neural Network (CNN) model for wildfire forecasting in Chapter 4. The model is designed to process static and dynamic variables by two separated branches, 2D CNN for static variables (e.g., land cover) and 3D CNN for dynamic variables that change over time (e.g., temperature). In addition to this contribution, we introduce the Location-aware Adaptive Normalization module (LOAN). Think of rainfall which is a dynamic variable. Heavy rainfall would cause different impacts depending on the landscape, e.g., the soil type. The main idea of LOAN is to learn an adaptive function to adaptively normalize or modulate dynamic features inside the network conditioned on the corresponding location attributes. The module can be

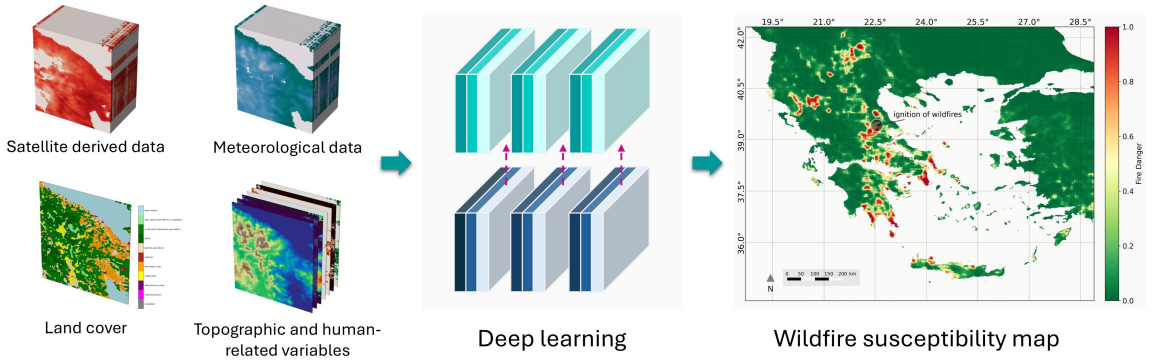


Figure 1.1: An example of wildfire forecasting with deep learning. The deep learning model starts with an initial condition as input using, e.g., satellite products and static features and predicts for the next day the susceptibility map of wildfire danger. The black circles represent an ignition of a large wildfire on the corresponding day.

integrated into any existing neural network architectures with multi-modality input data (Fig. 1.1). By integrating the LOAN module into the 2D/3D CNN multi-branch model, we show an improvement compared to previous works on wildfire forecasting. We also integrated LOAN for a model developed for flood forecasting in Chapter 7 and observe performance improvement for flood forecasting.

1.3.2 Focal-TSMP: Deep Learning for Vegetation Health Prediction and Agricultural Drought Assessment from a Regional Climate Simulation

Our second contribution is the development of the Focal-TSMP model in Chapter 5. Focal-TSMP is a deep learning model designed to address the challenge of long-term forecasting of vegetation health and extreme agricultural droughts. We use Focal-TSMP to predict satellite images conditioned on climate scenarios. As deep learning models have difficulties for forecasting satellite images on a long-term, Focal-TSMP synergizes regional climate simulation and deep learning. It leverages first the capability of the Terrestrial Systems Modeling Platform (TSMP) climate simulation to generate physically consistent long-term simulations as predictors. Then, it uses deep learning to map the simulation forecast variables directly into satellite-derived images like Normalized Difference Vegetation Index (NDVI) and Brightness Temperature (BT) as target. The predicted images are then used to derive different vegetation and agricultural drought indices like Vegetation Health Index (VHI). Figure 1.2 illustrates the framework of Focal-TSMP. The core blocks in Focal-TSMP are built on focal modulation networks, which are efficient alternatives to Transformers. Our evaluation demonstrates that Focal-TSMP produces predictions from TSMP data that align with actual satellite observations even with longer forecasting time, e.g., 2-3 years forecasts.

1.3.3 Identifying Spatio-Temporal Drivers of Extreme Events

In our Focal-TSMP work (Section 1.3.2), we focus on forecasting vegetation health and extreme agricultural droughts. The next contribution is to identify the drivers of extreme agricultural droughts. We therefore, introduce the new task of identifying spatio-temporal drivers of extreme events impacts in Chapter 6. This task is very challenging, because the drivers or anomalies may occur earlier in time and at a different location than the observed impacts, and the time delays between the impacts and

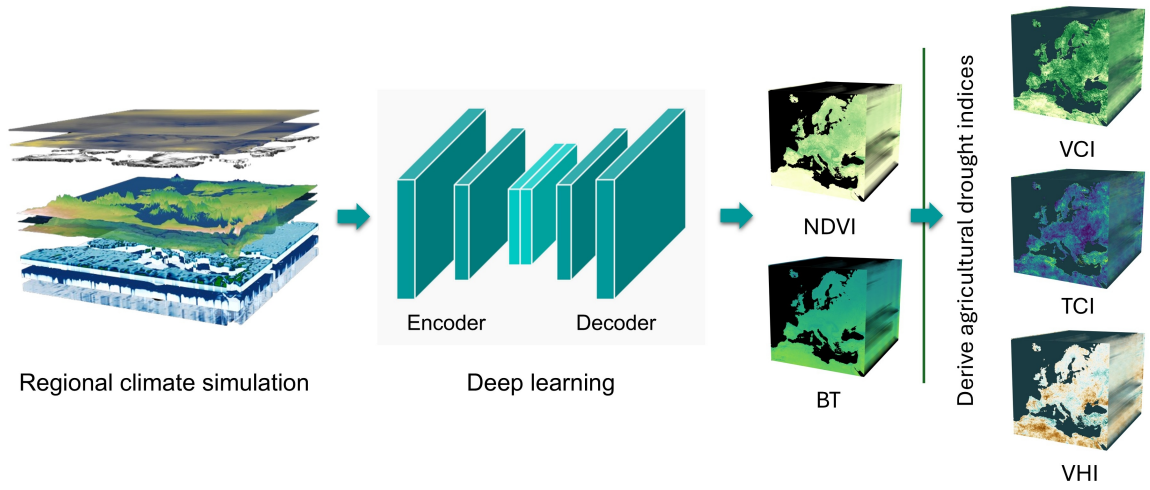


Figure 1.2: An overview of the proposed task with Focal-TSMP model in Chapter 5. First a regional climate simulation is generated by the Terrestrial Systems Modeling Platform (TSMP). The simulation could be a control run, e.g., a realization of the terrestrial hydrologic and energy cycles or reanalysis or it could be a land- or water use scenario based on human intervention. Then, a deep learning model takes the simulated land and atmospheric variables and predicts satellite images that are not part of the simulation output like Normalized Difference Vegetation Index (NDVI) and Brightness Temperature (BT). The predicted images and climatology are then used to derive vegetation health and agricultural drought indices like the Vegetation Condition Index (VCI), Thermal Condition Index (TCI), and Vegetation Health Index (VHI).

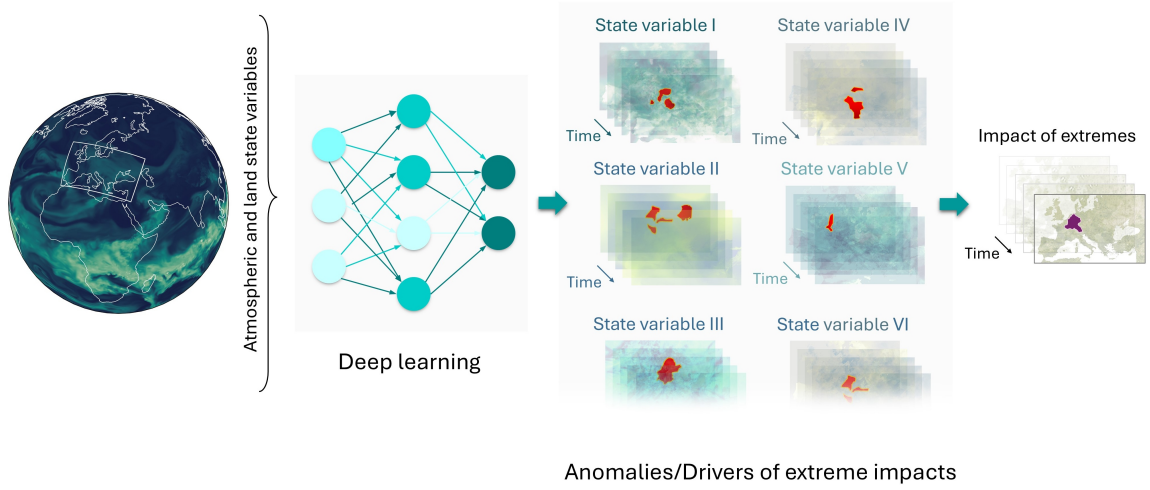


Figure 1.3: An overview of the task of identifying spatio-temporal relations between the measurable impacts of extremes like the vegetation health index ■ and their drivers like anomalies in state variables of the land-atmosphere and hydrological cycle ■. Anomalies or drivers may occur earlier in time and at a different location than the observed impact. This makes the task very challenging.

their drivers are dynamic (Fig. 1.3). For this, we propose a novel deep learning approach for identifying spatio-temporal drivers in climate data (e.g., anomalies in precipitation and soil moisture) that are

spatio-temporally correlated with the impacts of extreme events (e.g., extreme agricultural drought). The proposed model makes use of an advanced quantization technique which maps the input state variables into a binary representation (driver or non-driver). The model then predicts the impacts of extremes from the spatio-temporal binary representation maps. In this way, the model identifies only the drivers in the input variables that are spatio-temporally associated with the impacts of extremes. We further verify our approach on two long-term real-world reanalysis datasets including various physical variables from five biogeographical diverse regions. Another contribution factor is the absence of established benchmarks for evaluating this highly important task. As we have discussed in Section 1.2.5, annotations of impacts of extreme events can be derived from remote sensing observations or reported data. However, annotations of drivers or anomalies in the input variables do not exist. To address this gap, we propose a framework to generate synthetic datasets that mimic Earth observations properties including ground truth for drivers and anomalies. The framework can be used to generate synthetic datasets with different properties and relations between the impacts and drivers. Thus, synthetic datasets generated by this framework can serve as benchmarks to perform the quantitative evaluation of different approaches.

1.3.4 RiverMamba: A State Space Model for Global River Discharge and Flood Forecasting

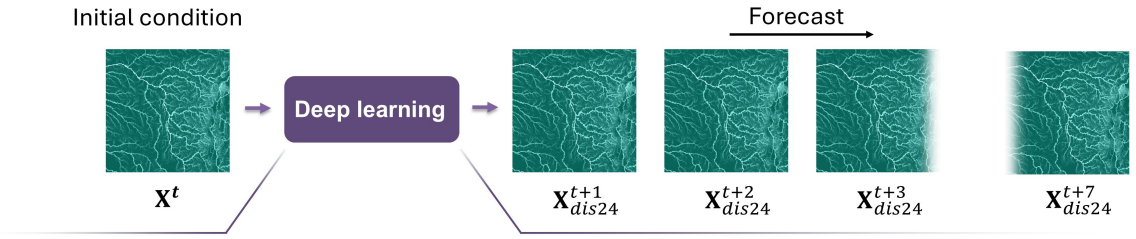


Figure 1.4: An overview of the task of medium-range river discharge and flood forecasting with deep learning. Ideally, the deep learning model should be a vision-like model that considers the spatial relations between points at and near bodies of water. This is different to the catchment concept or lumped approach, which does not consider spatial connectivity. In addition, the model should ideally consider uncertainty in the meteorological forcing.

Our final contribution is related to river discharge and flood forecasting (Fig. 1.4). In contrast to typical catchment concept or lumped approaches, we frame the task of flood forecasting as a vision task. Based on this, we introduce in Chapter 7 a novel methodology to hydrology called RiverMamba. We build RiverMamba such that it is based on efficient spatio-temporal blocks that enable the model to capture large-scale river networks and enhance the forecast for longer lead times. RiverMamba is able to integrate sparse gauged observations, river attributes, high-resolution reanalysis data, and weather forecast. Remarkably, this contribution results of the first deep learning approach that is able to provide maps of global river discharge and floods forecasting at 0.05° up to 7 days lead time, which is highly relevant to early warning. We evaluate RiverMamba on both long-term real-world reanalysis and observational data. We show that RiverMamba outperforms state of the art models like Google's LSTM and the physics-based Global Flood Awareness System (GloFAS) operated by the European Centre for Medium-Range Weather Forecasts (ECMWF) for global flood forecasting.

1.4 Dissertation Structure

The rest of this dissertation is organized as follows:

Chapter 2 provides preliminaries and background information that are related to the approaches shown in Chapters 4-7. The chapter starts with a compact overview of artificial neural network architectures and the core concepts behind them. Then, it introduces the datasets used in this dissertation and the data pre-processing necessary to run the experiments. In the end, the chapter describes the evaluation metrics that are adapted to assess the performance of the proposed approaches.

Chapter 3 provides a brief literature review of the related works for topics like wildfire, agricultural drought, and flood forecasting. It provides also an overview of the related works for the task of anomaly detection and identifying drivers of extreme events impacts. Finally, it includes a brief literature review for techniques related to the designing of neural networks architectures.

Chapter 4 introduces the LOAN module for adaptive normalization. It also introduces a multi-branch 2D/3D CNN deep learning model for wildfire forecasting. This chapter is based on *Shams Eddin et al.* (2023).

Chapter 5 introduces Focal-TSMP, a deep learning model that combines regional climate simulation with deep learning to forecast agricultural droughts. This chapter is based on *Shams Eddin and Gall* (2024a).

Chapter 6 introduces the task of identifying spatio-temporal drivers of extreme events impacts. It also introduces a framework to generate synthetic data that can be used as benchmarks for the proposed task. The chapter then introduces a deep learning approach for identifying spatio-temporal drivers in climate data that are spatio-temporally correlated with the impacts of extreme events. The chapter is based on *Shams Eddin and Gall* (2024c).

Chapter 7 addresses the task of forecasting river discharge and floods at global scale. In this chapter, RiverMamba is introduced and applied for flood forecasting. The chapter is based on *Shams Eddin et al.* (2025a).

Finally, conclusions and outlook are given in **Chapter 8** along with potential future directions.

CHAPTER 2

Background

We will begin this chapter by discussing basic artificial neural networks and their core concepts. We then move to more advanced architectures that are relevant to our work. Second, we introduce the datasets used in this dissertation. Third, we give a brief background regarding return periods and the definition of the scientific regions used in this dissertation. In the end, we describe the evaluation metrics used for the experiments.

Contents

2.1 Artificial Neural Network (ANN) . . . 12
2.1.1 Multi Layer Perceptron (MLP) . . . 12
2.1.2 Convolutional Neural Network (CNN) . . . 13
2.1.3 Long Short-Term Memory (LSTM) . . . 14
2.1.4 Convolutional Long Short-Term Memory (ConvLSTM) . . . 15
2.1.5 Vision Transformer (ViT) . . . 15
2.1.6 Focal Modulation Network . . . 17
2.1.7 State Space Model (SSM) and the Mamba Family . . . 18
2.2 Datasets . . . 19
2.2.1 FireCube . . . 19
2.2.2 Terrestrial System Modelling Platform (TSMP) . . . 21
2.2.3 Observational Remote Sensing Data . . . 23
2.2.4 ERA5-Land Reanalysis Data . . . 27
2.2.5 CERRA Reanalysis Data . . . 28
2.2.6 GloFAS . . . 30
2.2.7 GRDC Observational River Discharge Data . . . 32
2.2.8 HRES Meteorological Forecast . . . 33
2.2.9 CPC Precipitation Data . . . 34
2.2.10 LISFLOOD Static Data . . . 35
2.2.11 HydroRIVERS Static Data . . . 35
2.2.12 Diagnostic River Points for Flood Forecasting . . . 36
2.3 Return Periods and Flood Definition . . . 36
2.4 PRUDENCE Scientific Regions . . . 38
2.5 Evaluation Metrics . . . 39
2.5.1 Mean Absolute Error (MAE) . . . 39
2.5.2 Root Mean Square Error (RMSE) . . . 39
2.5.3 Bias . . . 39

2.5.4 Pearson Correlation Coefficient (R_p) . . . 39
2.5.5 Spearman Correlation Coefficient (R_s) . . . 40
2.5.6 Coefficient of Determination (R2) . . . 40
2.5.7 Kling-Gupta Efficiency (KGE) . . . 40
2.5.8 Precision . . . 40
2.5.9 Recall . . . 41
2.5.10 F1-score . . . 41
2.5.11 Intersection over Union (IoU) . . . 41
2.5.12 Overall Accuracy (OA) . . . 41
2.5.13 Area Under the Receiver Operating Characteristic Curve (AUROC) . . . 41

2.1 Artificial Neural Network (ANN)

Artificial neural networks (ANNs) (*Amari*, 1972) belong to the field of Machine Learning (ML). Although ANNs are inspired by the structure of human brain where neurons are connected with each other, we can think of an ANN as a learnable function f to map input x into output y via approximation and learnable weights. In general, an ANN consists of a stack of functions or layers with parameters and learnable weights. These layers are called the building blocks of an ANN and the ANN represented as a function $y = f(x)$ can be thought of as a composition of these layers or functions: $y = f(x) = f_M(f_{M-1}(...f_0(x)))$ for M layers or functions. ANNs start with random weights and are trained or optimized to minimize a predefined objective function $\mathcal{L}$. The common way to perform the optimization is to use gradient descent with backpropagation (*Linnainmaa*, 1976; *Rumelhart et al.*, 1986). In backpropagation, the gradients of $\mathcal{L}$ is computed with respect to the learnable weights and the weights are then updated iteratively to reach optimally a local minima of $\mathcal{L}$. In recent years, ANNs gained their popularity with the emerge of deep learning (DL) (*Schmidhuber*, 2015; *LeCun et al.*, 2015). In DL, the neural networks are constructed by stacking many layers (deep networks) and then these networks are trained with large-scale datasets to solve sophisticated tasks like image recognition and generation. This is different to typical ML models, since DL does not rely on feature engineering but rather learns to automatically extract complex patterns from the data and solve the task in an end-to-end manner.

In the next sections, we give an overview of common ANN architectures and their building blocks that are relevant to the methods developed in this dissertation.

2.1.1 Multi Layer Perceptron (MLP)

MLP is one of the simplest form of ANNs (*Rosenblatt*, 1958). MLP can be built by stacking multi linear or fully connected layers. A fully connected layer serves as an affine transformation of the input $\mathbf{X} \in \mathbb{R}^{C_i}$:

$$y = xW + b\,, \tag{2.1}$$

where $W \in \mathbb{R}^{C_i \times C_o}$ is the learnable weights of the linear transformation, $b \in \mathbb{R}^{C_o}$ is the translation or bias of the transformation, C_i is the input dimension, C_o is the out dimension of the layer, and

$y \in \mathbb{R}^{C_o}$ is the transformed output. Usually, an activation function (σ) is added to the output of the linear layer to add non-linearity. Typical activation functions are Rectified Linear Unit (ReLU) (*Fukushima*, 1969), Sigmoid, and Tanh functions. The output of each layer is then fed into the next layers sequentially (hence, MLP is also a type of feed-forward layers). Linear layers and MLP are very common in ANNs. Usually, they are used in a combination with other types of ANNs to build more advanced networks.

2.1.2 Convolutional Neural Network (CNN)

The basic form of CNNs was first introduced by *Fukushima* (1979). CNNs are especially designed to handle spatial data like images. Each CNN layer consists of fixed sized convolutional kernels or filters with learnable weights and bias. The filter is convolved with the input data with a receptive field bounded by the filter size, thus CNN preserves the receptive field over spatial connectivity. The learnable parameters of the filters are shared across the spatial input. Consequently, CNN have much better generalization and less number of parameters than MLP when it is applied to 2D tasks. By stacking CNN layers, the network learns to extract hierarchical feature representation from spatial data, e.g., going from low level features like edges and textures to high level features like semantics. CNNs are also combined with other techniques like pooling, padding and strides. Pooling is used to reduce the spatial resolution of the input sequentially, which enables the hierarchical feature extraction and reduces computational costs. Padding adds extra pixels (e.g., with values 0) to the boundaries of the image depending on the filter size to maintain convolution operators on the boundaries of the images. The stride defines the distance between two consecutive convolutional operators or it defines how the filter slides while it scans the input image. Sometimes, strides are also used for computational efficiency, e.g., to downsample an image by skipping some in-between locations. Usually the output of the convolutional layer is transformed with a non-linear activation function. For a 2D CNN, given an input image $x \in \mathbb{R}^{H_i \times W_i \times C_i}$ with C_i channels and $H_i \times W_i$ spatial dimension, the output of the convolutional layer with 2D filters of size (k_h, k_w) and C_o output channels will be $y \in \mathbb{R}^{H_o \times W_o \times C_o}$ where each element of y can be computed with the following convolutional operator '$*$':

$$y(h_o, w_o, c_o) = (x * W)(h_o, w_o, c_o) = \sum_{c_i=1}^{C_i} \sum_{i=-\lfloor k_h/2 \rfloor}^{\lfloor k_h/2 \rfloor} \sum_{j=-\lfloor k_w/2 \rfloor}^{\lfloor k_w/2 \rfloor} x[h_i - i, w_i - j, c_i] W[i, j, c_i, c_o] + b[c_o]\,, \quad (2.2)$$

where $W \in \mathbb{R}^{k_h, k_w, C_i, C_o}$ are the learnable filters and $b \in \mathbb{R}^{C_o}$ are the learnable biases.

Similarly, 3D CNN with a 3D filter (k_h, k_w, k_t) for spatio-temporal data can be formulated as follows:

$$\begin{aligned} y(h_o, w_o, t_o, c_o) &= (x * W)(h_o, w_o, t_o, c_o) \\ &= \sum_{c_i=1}^{C_i} \sum_{i=-\lfloor k_h/2 \rfloor}^{\lfloor k_h/2 \rfloor} \sum_{j=-\lfloor k_w/2 \rfloor}^{\lfloor k_w/2 \rfloor} \sum_{k=-\lfloor k_t/2 \rfloor}^{\lfloor k_t/2 \rfloor} x[h_i - i, w_i - j, t_i - k, c_i] W[i, j, k, c_i, c_o] + b[c_o]\,. \end{aligned} \quad (2.3)$$

There are also depth-wise convolutions where each input channel is convolved with one filter channel to reduce computations. CNNs have become a cornerstone for image and video processing applications, e.g., with early works like handwritten digits recognition (*LeCun et al.*, 1989; *Lecun et al.*,

1998), image classification (*Krizhevsky et al.*, 2012; *Simonyan and Zisserman*, 2015), object detection (*Ren et al.*, 2015), and video analysis tasks (*Tran et al.*, 2015; *Carreira and Zisserman*, 2017).

2.1.3 Long Short-Term Memory (LSTM)

LSTM belongs to a type of neural networks called Recurrent Artificial Neural Networks (RNNs) which are designed to handle sequential or temporal data (*Elman*, 1990). The basic building block of an RNN is the recurrent layer or unit which computes the hidden state h_t at time step t based on a linear combination of the input x_t at time step t and the previous hidden state h_{t-1} at time step $t-1$:

$$h_t = \sigma_h(h_{t-1}W_h + x_tW_x + b_h), \tag{2.4}$$
$$y_t = \sigma_y(h_tW_y + b_y)\,, \tag{2.5}$$

where W_x and W_h are the weights that control the contribution from the current input x_t and the previous hidden state h_{t-1}, respectively, σ_* is an activation function, weights W_y control how much information should be propagated into the current output y_t, and b_h and b_y are learnable bias biases. The initial hidden state at time step t_0 is usually set heuristically, e.g., $h_{t=0} = 0$. By stacking multiple recurrent layers, the final output $y_{t=T}$ at last time step $t = T$ can be obtained as a function of all inputs $X_{1:T-1} = (x_1, \ldots, x_{T-1})$ from the previous time steps. Like this, RNNs transform the input sequence $X_{1:T} = (x_1, \ldots, x_T)$ into a new sequence $Y_{1:T} = (y_1, \ldots, y_T)$ (sequence-to-sequence modeling). This gives RNNs a memory mechanism, very useful to perform long-term modeling along a sequence. To handle variable length of X, the weights W_h, W_x, W_y, b_h, b_y are shared across the time steps and the RNN is applied recursively across time.

However, this neural network formulation has been shown to impact the stability of RNNs training due to vanishing or exploding gradients when handling long sequences (*Hochreiter*, 1991; *Bengio et al.*, 1994). Thus, LSTM (*Hochreiter and Schmidhuber*, 1997) as a variant of RNN has been proposed to improve the ability of RNN to learn effectively from long sequence. Following *Graves* (2013), the formulation of an LSTM unit can be written as follows:

$$i_t = \sigma(h_{t-1}W_{hi} + x_tW_{xi} + c_{t-1}W_{ci} + b_i), \tag{2.6}$$
$$f_t = \sigma(h_{t-1}W_{hf} + x_tW_{xf} + c_{t-1}W_{cf} + b_f), \tag{2.7}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(h_{t-1}W_{hc} + x_tW_{xc} + b_c), \tag{2.8}$$
$$o_t = \sigma(h_{t-1}W_{ho} + x_tW_{xo} + c_tW_{co} + b_o), \tag{2.9}$$
$$h_t = o_t \odot \tanh(c_t)\,, \tag{2.10}$$

where $i_t, f_t,$ and o_t are the input, forget, and output gates, respectively, c_t acts like a memory cell to accumulate the state information, σ is a sigmoid activation function, $\odot$ denotes the Hadamard product, $W_{hi}, W_{xi}, W_{ci}, W_{hf}, W_{xf}, W_{cf}, W_{hc}, W_{xc}, W_{ho}, W_{co},$ and W_{co} are the learnable weights, and $b_i, b_f, b_c,$ and b_o are the learnable biases.

The input gate i_t, controls the information flow from the input to update the memory cell. The forget gate f_t controls the information that should be kept in the memory cell at each time step. Finally, the output gate o_t determines the output depending on the previous hidden state, current input, and current state. The biggest advantage here is that the gates and memory cell can control the flow of the gradient that should be kept for an LSTM cell at specific time step t, helpful with

the constant error carousels (*Hochreiter and Schmidhuber*, 1997). In contrast to typical RNNs, this formulation allows the LSTM to handle longer sequences more efficiently.

Unfortunately, when training an LSTM, the backpropagation through time and the recursion formulation still lead to high computational cost, memory consumption, and limited parallelization. To address these issues, alternative of LSTM have been proposed in recent years including Transformer (*Vaswani et al.*, 2017), xLSTM (*Beck et al.*, 2024), and Mamba (*Gu and Dao*, 2024).

2.1.4 Convolutional Long Short-Term Memory (ConvLSTM)

Since LSTM (Section 2.1.3) does not consider spatial correlations and 3D CNN (Section 2.1.2) becomes computationally expensive with long sequences, building a model that combines both is intuitive. ConvLSTM was first introduced by *Shi et al.* (2015) for precipitation nowcasting. It extends the LSTM by combining the benefits of both LSTM and CNN for spatio-temporal modeling. We can think of ConvLSTM similar to LSTM, but with convolutional operations enabling additional spatial modeling. The structure is similar, but the input, gates, hidden states, and output are all represented as 3D tensors. ConvLSTM encodes spatio-temporal information using the following formulation:

$$i_t = \sigma(h_{t-1} * W_{hi} + x_t * W_{xi} + c_{t-1} \odot W_{ci} + b_i), \tag{2.11}$$
$$f_t = \sigma(h_{t-1} * W_{hf} + x_t * W_{xf} + c_{t-1} \odot W_{cf} + b_f), \tag{2.12}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tanh(h_{t-1} * W_{hc} + x_t * W_{xc} + b_c), \tag{2.13}$$
$$o_t = \sigma(h_{t-1} * W_{ho} + x_t * W_{xo} + c_t \odot W_{co} + b_o), \tag{2.14}$$
$$h_t = o_t \odot \tanh(c_t)\,, \tag{2.15}$$

where '$*$' denotes the convolutinoal operator and $\odot$ is the Hadamard product. Here, each future state on a 2D grid will be determined based on its current input and previous states of its local neighborhood.

2.1.5 Vision Transformer (ViT)

The transformer model we discuss here is based on the work of *Vaswani et al.* (2017). It was initially introduced for Natural language processing (NLP). From there on, transformers have stood the test of time and their application covered many tasks across various fields in science. The application of transformer in computer vision has led to the development of Vision Transformer (ViT). ViT is designed to handle 2D images and the core module of it is the self-attention mechanism. With self-attention, the neural network can capture how each token (e.g., pixel) relates to every other token in the input sequence. In contrast to CNNs, ViTs with self-attention modules can handle long-range interactions across tokens more skillfully. ViT is built upon four main concepts; (a) tokenization, (b) self-attention, (c) multi-head attention, and (d) positional encoding.

Tokenization. Since transformer operates originally on a sequence of words, applying it to an image requires us to convert the image into a 1D sequence. This is done via tokenization. In tokenization, the image is divided into non-overlapping tokens (also called patches) where each token could be 1 pixel or 2×2 pixels or even more. These tokens are then flattened into a 1D sequence and usually a linear layer is applied on each token to perform the embedding of the lower dimensionality. This embeddings will be the input to the ViT.

Self-attention. It is used to capture the interaction between the tokens and to capture long-range dependence. In a nutshell, the self-attention module aims to transfer pixel representations of a given image into a new feature representation based on a weighted aggregation of interactions between every individual token and its surrounding. This mechanism allows the neural network to focus more on relevant regions of the input image. For a 1D sequence $x \in \mathbb{R}^{S \times C_i}$, where S is the number of tokens and C_i is the input dimension, the tokens are projected linearly into queries $\mathcal{Q}$, keys $\mathcal{K}$, and values $\mathcal{V}$ using weights matrices $W_{\mathcal{Q}} \in \mathbb{R}^{C_i \times D_k}, W_{\mathcal{K}} \in \mathbb{R}^{C_i \times D_k}, W_{\mathcal{V}} \in \mathbb{R}^{C_i \times D_v}$. The self-attention can be computed with scaled dot-product as follows:

$$\text{Attention}(\mathcal{Q}, \mathcal{K}, \mathcal{V}) = \text{softmax}\Big(\frac{\mathcal{Q}\mathcal{K}^{\text{T}}}{\sqrt{D_k}}\Big)\mathcal{V}\,. \tag{2.16}$$

Self-attention computes the dot-product between queries and keys and applies softmax to obtain an attention matrix or weights of a linear combination of the input tokens. The weights is normalized by an expected variance D_k to counteract the effect of exploding gradient with large values of D_k (*Vaswani et al.*, 2017). ViT uses a normalization layer (*Ba et al.*, 2016) before the self-attention and it incorporates an MLPs as feed-forward layers to further process the output of the self-attention. The feed-forward layer with an activation function adds non-linearity and generates a new representation of each token using the computed contextualized representation from the preceding self-attention module. Similar to self-attention, there is also a cross-attention mechanism in transformers where $\mathcal{Q}$, $\mathcal{K}$, and $\mathcal{V}$ come from different sources, e.g., $\mathcal{Q}$ will come from the input tokens and $\mathcal{K}$ and $\mathcal{V}$ come from a different source of input like conditioning tokens.

Multi-head attention. Instead of computing only one self-attention with a dimension D_k, multi-head attention performs the self-attention n times in parallel over chunks of queries, keys, and values. Each chunk is a subset of the original queries, keys, and values (e.g., with a dimensionality $\frac{D_k}{n}$). The output values from the attention heads are concatenated and transformed to generate the final output.

Positional encoding. In contrast to CNN which preserves spatial information by scanning the image with filters, ViT converts the image into 1D sequence via tokenization and since dot-product attention is agnostic about the position of the tokens, positional encoding is added to the tokens to preserve the relative spatial relations between the tokens. Positional encoding is usually added to the tokens via sinusoidal encoding (*Vaswani et al.*, 2017) after the tokenization.

ViT consists of transformer blocks. By stacking multiple blocks, the neural networks learn to capture complex relations between the tokens to solve complicated tasks. Transformers allow parallelization and better gradient flow during training (e.g., there is no recursion), solving a critical limitation in LSTM (Section 2.1.3) and ConvLSTM (Section 2.1.4), especially with long sequences. The network design of ViT along with the multi-head self-attention mechanism allow ViTs to stand as the state-of-the-art backbone in recent DL models.

Despite this powerful transforming process, the computational requirement of a standard ViT has been a limitation when applying it to vision tasks, especially with high resolution images. This is because the self-attention has quadratic computational complexity with sequence length $\mathcal{O}(S^2)$ in memory footprint and time.

To overcome this issue, Swin Transformer (*Liu et al.*, 2021, 2022a) divide the image into windows and applies attention within them. The windows are shifted as the input progresses through the transformer blocks to introduce interactions between the pixels on a global scale. In addition, Swin Transformer builds the model upon a hierarchical feature representation by merging the

tokens (down-sampling) in the deeper level of the network. Considering that each shifted window in Swin Transformer has $(M \times M)$ tokens, the computational complexity of Swin Transformer is $\mathcal{O}(4HWC^2 + 2M^2HWC)$ (*Liu et al.*, 2021) in comparison to the vanilla ViT with $\mathcal{O}(4HWC^2 + 2(HW)^2C)$, where C is the channel dimension and H and W are the height and width fo the image, respectively.

Flash Attention (*Dao et al.*, 2022; *Dao*, 2024) reduces the complexity using hardware optimization techniques. It scales the memory linearly with the input tokens. Flash attention divides the computation of the attention matrix into smaller blocks. These blocks are loaded to the fast memory of the Graphics Processing Unit (GPU) called SRAM to compute the attention output and the final results are then combined together. During backpropagation, the gradient is recomputed using the stored output and the softmax normalization, which is more efficient than reading the large intermediate attention matrix from the High Bandwidth Memory (HBM) of the GPU. Kernel fusion is also used in Flash attention to accelerate the memory-bound operations in transformers.

Building upon the successes of ViT in 2D image processing, transformers have been also applied to spatio-temporal tasks. For example, TimeSformer was introduced in *Bertasius et al.* (2021) and Swin Transformer was extended to Video Swin Transformer in *Liu et al.* (2022b) for video understating.

2.1.6 Focal Modulation Network

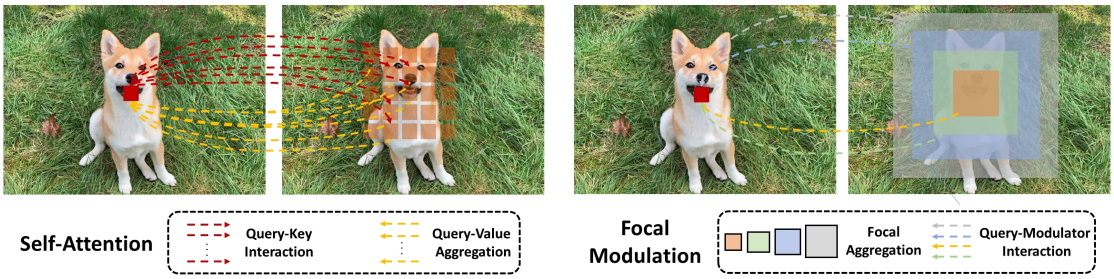


Figure 2.1: Illustrative comparison between window-based self-attention (Section 2.1.5) and focal modulation. Self-attention starts with query-key interactions and then performs query-value aggregation. Focal modulation aggregates contextual information around each query and ends with interactions to modulate the query with the aggregated contextual information. Figure adapted from *Yang et al.* (2022).

Focal modulation networks (*Yang et al.*, 2022) have been introduced to substitute the self-attention mechanism (Section 2.1.5) with a lightweight focal module. Self-attention starts with interaction (query-key interaction) and then aggregates the information of interactions (query-value aggregation). In contrast, focal modulation starts with contextual aggregation using depth-wise CNN layers around each query and ends with interactions to modulate the query with the aggregated contextual information. An overview of focal modulation is shown in Fig. 2.1. The design of a focal modulation network mimics the human eye where the focal region to focus on can be changed adaptively via a learned CNN. The focusing operator is implemented as a lightweight element-wise multiplication which allows each pixel or token to interact with its neighborhood. As shown in Fig. 2.2,

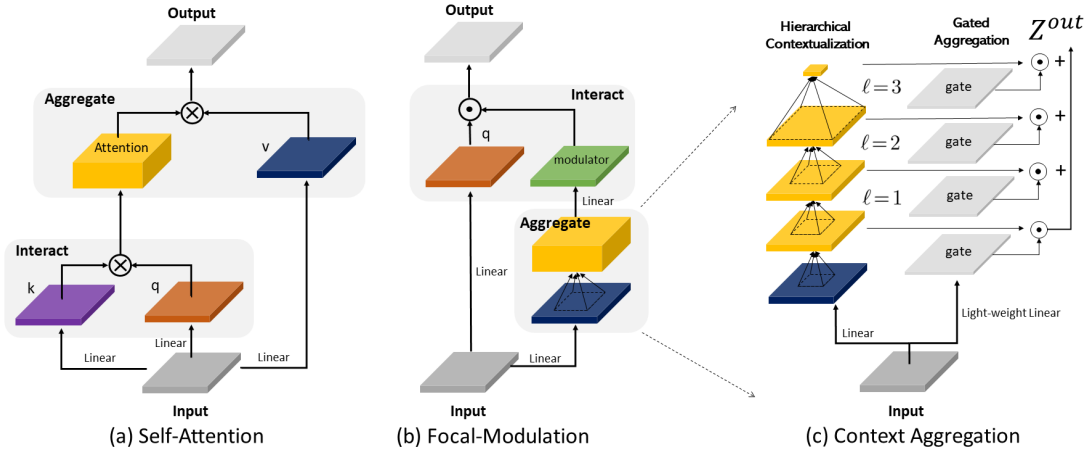


Figure 2.2: Comparison between self-attention (a) (Section 2.1.5) and focal modulation (b). The context aggregation in focal modulation is illustrated on the right side (c). Figure adapted from *Yang et al.* (2022).

the focal modulation mechanism consists of three main parts; (a) hierarchical contextualization, (b) gated aggregation, and (c) interaction.

Hierarchical contextualization. It is based on Focal Transformer (*Yang et al.*, 2021a) and it extracts features from local to global with different granularity and shared depth-wise convolution. This reduced the computations compared to self-attention which requires a summation or aggregation over non-shareable operator.

Gated aggregation. The gates control the information flow and condense the aggregated contexts into a modulator.

Interaction. In the interaction step, the constructed modulator interacts with the input (represents the query) using a lightweight element-wise operator to have the final output.

This overall design makes focal modulation more efficient for visual token interaction than self-attention according to *Yang et al.* (2022). Focal modulation has four main linear projections along with a lightweight linear gating layer and a stack of L focal levels with depth-wise convolutions for the hierarchical contextualization (Fig. 2.2). Hence, the complexity of focal modulation is $\mathcal{O}(4HWC^2 + HWC(2L+3) + HWC\sum_l (r_l)^2)$ (*Yang et al.*, 2022) in comparison to Swin Transformer with $\mathcal{O}(4HWC^2 + 2M^2HWC)$, where r_l is the kernel size at level l.

Based on focal modulation mechanism, DL models were developed for medical image segmentation (*Naderi et al.*, 2022; *Rasoulian et al.*, 2023), change detection for remote sensing data (*Fazry et al.*, 2023), and video action recognition for 3D data (*Wasim et al.*, 2023).

2.1.7 State Space Model (SSM) and the Mamba Family

Linear State Space Models (SSMs) (*Gu et al.*, 2021) and structured SSMs like (*Gu et al.*, 2022) and (*Smith et al.*, 2023) were primarily introduced for long-sequence modeling in NLP. Recently, Mamba (*Gu and Dao*, 2024) introduced the selective scan mechanism, enabling efficient training and linear-time inference. The Mamba block is based on a state-space model that transforms a 1D sequence of states $x(t)$ into another representation $y(t)$ through an implicit hidden latent state $h(t)$

and a first-order ordinary differential equation:

$$h'(t) = \mathbf{A}h(t) + \mathbf{B}x(t)\,, \quad y(t) = \mathbf{C}h(t) + \mathbf{D}x(t)\,. \tag{2.17}$$

A works as a forget gate for the hidden states, **B** determines the extent of the contribution from the current state $x(t)$, **C** is a gate to control how much information should be propagated into the output $y(t)$, **D** can be viewed as a skip connection of the state $x(t)$, and y_t is the transformed representation.

To integrate Eq. (2.17) into a deep learning framework, S4 introduced by *Gu et al.* (2022) parametrized the system with the matrices $(\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D})$ and discretized it with a timescale parameter Δ:

$$h_t = \bar{\mathbf{A}}h_{t-1} + \bar{\mathbf{B}}x_t\,, \quad y_t = \mathbf{C}h_t + \mathbf{D}x_t\,, \tag{2.18}$$

$$\bar{\mathbf{A}} = e^{(\Delta\mathbf{A})}\,, \quad \bar{\mathbf{B}} = (\Delta\mathbf{A})^{-1}(e^{(\Delta\mathbf{A})} - \mathbf{I})\Delta\mathbf{B}\,, \tag{2.19}$$

where $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ are the discretized versions of the system. In a follow up work, *Smith et al.* (2023) (S5) added the parallel scans which made the SSMs more efficient on sequence-to-sequence transformation. Parallel scanning is a hardware-aware implementation of the SSM on the GPU to avoid the sequential recurrence. Recently, *Gu and Dao* (2024) (S6) proposed to make Eqs. (2.18) and (2.19) time-variant. To this end, the parameters $\mathbf{B}(x)$, $\mathbf{C}(x)$, and $\Delta(x)$ become dependent on the input state x. This representation of a state-space model is called Mamba, which is an efficient alternative to transformers (Section 2.1.5), particularly when processing high resolution images or long sequences of data.

In comparison to ViT, Mamba's computational complexity scales linearly with the input tokens $\mathcal{O}(S)$. In addition, the transformer architecture does not perform a state tracking, meaning that it might suffer with counting tasks. In comparison to LSTM, Mamba utilizes a gated linear RNN architecture, conceptually similar to an LSTM with linear activations. However, what truly sets Mamba apart is the hardware-aware implementation of a parallel scan that addresses the inherent sequential bottleneck of traditional RNNs. Moreover, LSTM can be thought of as a sequential and gate-controlled RNN. It models the dynamic of the state with input, forget, and output gates. These gates are controlled by a fixed set of weights at every time step. In contrast, the state update in Mamba with the core parameters $\mathbf{B}(x)$, $\mathbf{C}(x)$, and $\Delta(x)$ is selective and input-dependent.

2.2 Datasets

2.2.1 FireCube

There are only very few publicly available datasets for wildfire forecasting and they differ significantly in the observational variables, the spatial and temporal resolutions, and the task that needs to be addressed. A related dataset is the Next Day Wildfire Spread dataset (*Huot et al.*, 2022) where the task is to predict wildfire spread. It is formulated as a scene classification task and not as a pixel-wise wildfire forecasting task as it is proposed in the FireCube dataset (*Prapas et al.*, 2022b) and addressed in this dissertation. The FireCube dataset was first published in *Prapas et al.* (2021) and extended later in *Kondylatos et al.* (2022). It includes multivariate spatio-temporal data streams with 90 variables from the years 2009-2021 with a resolution of 1 km $\times$ 1 km $\times$ 1 day. The area is 1253 km $\times$ 983 km, covering parts of the Eastern Mediterranean.

Variable name	Long name	Unit	Reference
DEM	digital elevation model	m	*Bashfield and Keim* (2011)
SLOPE	slope	◦	*Bashfield and Keim* (2011)
ROAD_DISTANCE	distance to roads	m	*Tatem* (2017)
WATERWAY_DISTANCE	distance to waterway	m	*Tatem* (2017)
POP_DENS	population density	humans/km^2	*Tatem* (2017)
CLC	Copernicus Corine land cover fractions of 10 classes	%	*Büttner* (2014)

Table 2.1: The static variables used in FireCube dataset.

Variable name	Long name	Unit	Reference
lst_day	day land surface temperature	K	*Wan et al.* (2015)
lst_night	night land surface temperature	K	*Wan et al.* (2015)
NDVI	normalized difference vegetation index	unitless	*Didan* (2015)
sminx	soil moisture index	unitless	*Cammalleri et al.* (2017)
max_t2m	maximum 2m temperature	K	*Muñoz Sabater et al.* (2021)
max_wind_speed	maximum wind speed	m/s	*Muñoz Sabater et al.* (2021)
min_rh	minimum relative humidity	%	*Muñoz Sabater et al.* (2021)
max_tp	total precipitation	m	*Muñoz Sabater et al.* (2021)
max_d2m	maximum 2m dew point temperature	K	*Muñoz Sabater et al.* (2021)
max_sp	maximum surface pressure	Pa	*Muñoz Sabater et al.* (2021)

Table 2.2: The dynamic variables used in FireCube dataset.

The variables include meteorological data (*Muñoz Sabater et al.*, 2021), satellite-derived products (*Didan*, 2015; *Wan et al.*, 2015), topographic features (*Bashfield and Keim*, 2011), human-related activities (*Tatem*, 2017), and historical fire records (*San-Miguel-Ayanz et al.*, 2013; *Giglio et al.*, 2016). In addition, Copernicus Corine Land Cover (CLC) (*Büttner*, 2014) and Fire Weather Index (FWI) (*Steinfeld*, 2022) are provided. The target is to predict for each pixel if a wildfire will ignite and become large (> 0.3 km^2) in the next day. The task is equivalent to binary classification, where the positive class represents a wildfire event.

Since wildfire forecasting is essentially an imbalanced classification task, *Kondylatos et al.* (2022) extracted samples as follows: For a target day $T+1$, the static variables form a patch of 25 km $\times$ 25 km centered around the target pixel at day T. In contrast, the dynamic variables consist of 25 km $\times$ 25 km $\times$ 10 days time series of observations from days $T-9$ until T. For each positive sample, a few negative samples from different locations are sampled. Although the negatives are from different locations, they are sampled from regions that have a similar land cover distribution as the positive samples to make the task more difficult.

Overall the dataset includes 71471 samples for training (13518 positives and 57953 negatives for the years 2009-2018), 6430 samples for validation (1300 positives and 5130 negatives for the year 2019) and 42820 samples for testing (1228 positives and 4860 negatives for the year 2020 and 4407 positives and 32325 negatives for the year 2021). The year 2021 in the test set contains an extreme wildfire season in Greece (*Giannaros et al.*, 2022; *Kondylatos et al.*, 2022). The extracted samples are available from *Prapas et al.* (2022c).

In Chapter 4, we use from the described dataset the same variables as in *Kondylatos et al.* (2022). This includes 15 static variables (Table 2.1) and 10 dynamic variables (Table 2.2).

2.2.2 Terrestrial System Modelling Platform (TSMP)

2.2.2.1 Regional Earth System Simulation

We use the simulation produced by the Terrestrial System Modelling Platform version 1.1. (TSMP) at the Research Centre Jülich (FZJ) IBG-3 Institute and originally described in *Shrestha et al.* (2014) and *Gasper et al.* (2014). The simulation used in this dissertation is first introduced in *Furusho-Percot et al.* (2019). TSMP is a physics-based integrated simulation representing a realization of the terrestrial hydrologic and energy cycles that cannot be directly obtained from measurements. Its setup consists of three main interconnected model components:

- The Consortium for Small Scale Modelling (COSMO) version 5.01 is a numerical weather model to simulate the diabatic and adiabatic atmospheric processes (*Baldauf et al.*, 2011).
- The Community Land Model (CLM) version 3.5 is used to simulate the bio-geophysical processes on the land surface (*Oleson et al.*, 2004, 2008).
- ParFlow version 3.2. is a hydrological model to explicitly simulate the 3D dynamic processes of water in the land surface and underground (*Jones and Woodward*, 2001; *Kollet and Maxwell*, 2006; *Jefferson and Maxwell*, 2015; *Maxwell et al.*, 2015; *Kuffour et al.*, 2020).

The ECMWF ERA-Interim data (*Dee et al.*, 2011) were used to define the initial and boundary conditions for the simulation. Based on this setup, a spinup of 10 years (1979-1988) was conducted to initialize the surface and subsurface hydrologic and energy conditions and to reach the dynamic equilibrium with the atmosphere before the actual run (1989-2019). The three model components were fully coupled via the OASIS3 coupler (*Valcke*, 2013) to form a unified soil-vegetation-atmosphere model. This scheme was built without nudging allowing the free-running of the simulated variables. Thus, TSMP is ideal for representing the heterogeneity of the water cycle from the subsurface to the top atmosphere in a free evolution. In addition, the long-term simulation is performed for a historical time period from January 1989 until summer in September 2019 with output variables aggregated on a daily basis and extending over the Europe EURO-CORDEX EUR-11 domain (*Giorgi et al.*, 2009; *Gutowski Jr. et al.*, 2016; *Jacob et al.*, 2020). The grid specification for TSMP is a standardized rotated coordinate system ($\phi_{(rotated\ pole)} = 39.5^\circ$ N, $\lambda_{(rotated\ pole)} = 18^\circ$ E) with a spatial resolution of $\sim 0.11^\circ$ (~ 12.5 km) and 412×424 grid cells in the rotated latitudinal and longitudinal direction, respectively. These spatio-temporal dimensions and the model setup make TSMP suitable for climatological studies at a continental scale.

2.2.2.2 Preprocessing of TSMP Simulation

For experiments on TSMP in Chapter 5, we excluded the lateral boundary relaxation zone by removing invalid grid points from the boundaries. This results in a final grid with 397×409 grid cells in the latitudinal and longitudinal direction, respectively. We selected variables available within the period applicable for the analysis. This results in 29 main variables from COSMO, 8 variables from CLM, and 2 main variables from ParFlow. Additionally, we used 3 static variables from the analysis

Model	Variable name	Long name	Unit	Level
COSMO	awt	atmosphere total water content	kg m^{-2}	1
	capec	specific convectively available potential energy	j kg^{-1}	1
	capeml	cape of mean surface layer parcel	j kg^{-1}	1
	ceiling	cloud ceiling height (above mean sea level)	m	1
	cli	vertical integrated cloud ice	kg m^{-2}	1
	clt	total cloud fraction	1	1
	clw	vertical integrated cloud water	kg m^{-2}	1
	hudiv	atmosphere water divergence	kg m^{-2}	1
	hur2	2m relative humidity	%	1
	hur(200, 500, 850)	relative humidity (at 200, 500 and 850 hpa)	%	3
	hus2	2m specific humidity	1	1
	hur(200, 500, 850)	relative humidity (at 200, 500 and 850 hpa)	1	3
	incml	convective inhibition of mean surface layer parcel	j kg^{-1}	1
	pr	precipitation	kg m^{-2}	1
	prc	convective precipitation	kg m^{-2}	1
	prg	large scale precipitation	kg m^{-2}	1
	prt	total rain water content vertically integrated	kg m^{-2}	1
	ps	surface air pressure	pa	1
	psl	sea level pressure	pa	1
	snt	total snow content vertically integrated	kg m^{-2}	1
	ta(200, 500, 850)	air temperature (at 200, 500 and 850 hpa)	K	3
	tch	drag coefficient of heat	1	1
	td2	2m dew point temperature	K	1
	ua(200, 500, 850)	eastward wind (at 200, 500 and 850 hpa)	m s^{-1}	3
	uas	eastward near-surface wind velocity	m s^{-1}	1
	va(200, 500, 850)	northward wind (at 200, 500 and 850 hpa)	m s^{-1}	3
	vas	northward near-surface wind velocity	m s^{-1}	1
	zg(200, 500, 850)	geopotential height (at 200, 500 and 850 hpa)	m	3
	zmla	height of boundary layer	m	1
CLM	evspsbl	evapotranspiration	mm s^{-1}	1
	hfls	surface upward latent heat flux	w m^{-2}	1
	hfss	surface upward sensible heat flux	w m^{-2}	1
	prsn	snowfall flux	kg m^{-2} s^{-2}	1
	prso	precipitation on ground	kg m^{-2} s^{-2}	1
	rlds	incoming shortwave radiation	w m^{-2}	1
	tas	near-surface air temperature	K	1
	trspsbl	transpiration	w m^{-2}	1
ParFlow	sgw	groundwater saturation	1	15
	wtd	water table depth	m	1

Table 2.3: Technical details on the output variables in the TSMP EUR-11 simulation. For more information on the data, we refer to *Furusho-Percot et al.* (2019).

in *Poshyvailo-Strube et al.* (2022). In order to connect local-related characteristics to climate conditions, we computed 3 additional static variables from the static variables described in Table 2.4. We computed slope (*Horn*, 1981) and roughness (*Wilson et al.*, 2007) from orography and distance to water from the land/sea mask. Variable descriptions are listed in Tables 2.3 and 2.4. To avoid overfitting or the domination of few input variables, we normalized the input of TSMP by subtracting the mean and dividing by the standard deviation corresponding to each input variable. These statistics were computed only from the years that are used for training. The invalid values of pixels

Model	Variable name	Long name	Unit	Level
CLM	orog	surface height or digital elevation model (DEM)	m	1
	sftlf	land-sea fraction	%	1
	zbot	atmospheric reference height (from COSMO to CLM)	m	1
computed from land-sea fraction	-	distance to water	km	1
computed from orography	-	roughness	1	1
computed from orography	-	slope	○	1

Table 2.4: Technical details on the static variables from CLM in the TSMP EUR-11 simulation and the computed static variables.

were replaced with zero values as input to the DL model. Finally, we reserved the years 1989-2009 and 2013-2016 for training, 2010-2011 and 2017 for validation, and 2012, 2018-2019 for testing. The split is done due to the nature of remote sensing ground truth data which consist of two satellite systems (more details are in Section 2.2.3.2).

2.2.3 Observational Remote Sensing Data

Worldwide satellite observations allow for near real-time monitoring of drought and vegetation conditions. In practice, vegetation states can be estimated from land surface reflectances acquired from satellites. As a result, the reflectances on the ground can be employed as agricultural drought indicators and as proxies for vegetation health.

In this dissertation, satellite-based vegetation health products were obtained from the NOAA Center for Satellite Applications and Research (STAR), `https://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/index.php`. The blended version (*Yang et al.*, 2020) is composed of long-term remote sensing data derived from two systems of satellites: The Advanced Very High-Resolution Radiometer (AVHRR) from 1981 to 2012 and its successor Visible Infrared Imaging Radiometer Suite (VIIRS) from 2013 onward. The dataset includes two essential products, namely NDVI and BT (Table 2.5). NDVI is computed from the red (ρ_R) and near-infrared (ρ_{NIR}) bands:

$$\text{NDVI} = \frac{(\rho_{NIR} - \rho_R)}{(\rho_{NIR} + \rho_R)}. \tag{2.20}$$

The NDVI is unitless and given in the range $[-0.1, 1]$. Same NDVI values should not be interpreted similarly for different ecosystems. In other words, the interpretation is highly dependent on the location and ecosystem productivity (*Kogan*, 1995a). BT is derived from the infrared (ρ_{IR}) band and given in Kelvin (K) within the range $[0, 400]$. The dataset is provided globally with 0.05° (~ 4 km) spatial resolution mapped into the Plate Carrée projection. NOAA Vegetation Products (VP) have been broadly used for research and real-world applications. To handle high frequency noise caused by clouds, aerosol, and atmospheric variation along with different random error sources, NDVI and BT were temporally aggregated into smoothed noise reduced weekly products. In addition, post-launch calibration coefficients and solar/sensor zenith angles are applied to account for sensor degradation and orbital drift. The weekly temporal coverage is needed for outliers and discontinuities removal and is suitable to study the phenological phases of vegetation and consequently to define agricultural drought (*Yang et al.*, 2020; *Kogan et al.*, 2011) as we will see in the next section.

Satellite system	Spectral band	Spectral range (μm)
AVHRR	ρ_R	0.58 - 0.68
	ρ_{NIR}	0.725 - 1.1
	ρ_{IR}	10.3 - 11.3
VIIRS	ρ_R	0.600 - 0.680
	ρ_{NIR}	0.846 - 0.885
	ρ_{IR}	10.500 - 12.400

Table 2.5: Technical details on the spectral channel characteristics for Advanced Very High Resolution Radiometer (AVHRR) and Visible Infrared Imaging Radiometer Suite (VIIRS).

2.2.3.1 Satellite-derived Extreme Agricultural Drought

It is generally challenging to define what exactly constitutes an extreme. Extremes can be categorized from the perspective of their impacts. For instance, extreme drought can be categorized into 4 types based on their impacts (*Liu et al.*, 2016a; *Hao et al.*, 2018); meteorological or climatological drought, agricultural drought (*Meza et al.*, 2020), hydrological drought, and socioeconomic drought (*Pedro-Monzonís et al.*, 2015). Meteorological drought is mainly related to the dryness and can be defined based on a deficiency in temperature or precipitation. Agricultural drought measures the impact of stress on vegetation and usually defined as soil water deficits. It is also widely conceived that drought originates and progresses from meteorological into agricultural drought (*Huang et al.*, 2015; *Zhang et al.*, 2017). Hydrological drought is related to the water storage. While socioeconomic drought can be measured based on supply and demand related to weather and deficit in water supply. In this dissertation, we are interested in extreme agricultural droughts.

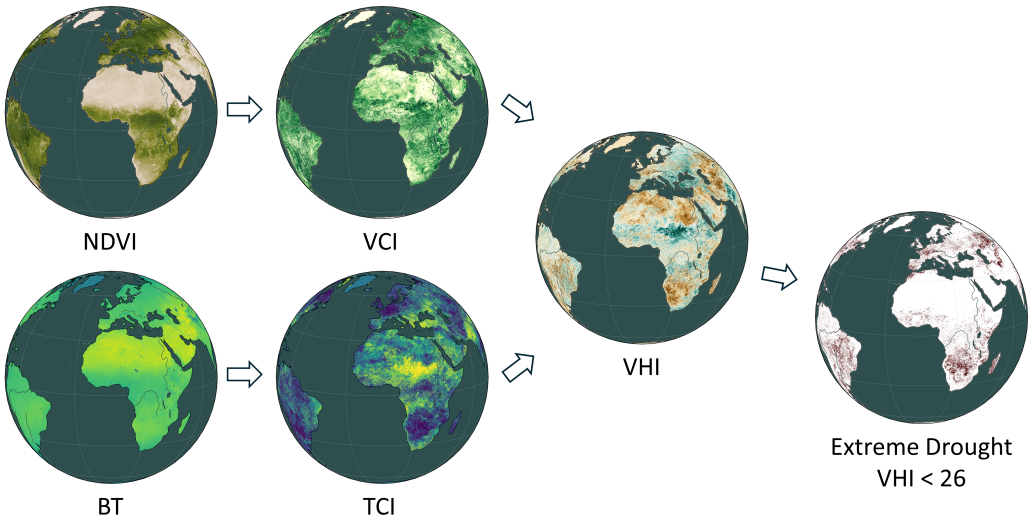


Figure 2.3: An overview of the extreme agricultural droughts definition from remote sensing.

Based on NDVI, BT and their long-term climatologies, the upper and lower bounds of the ecosystem (minimum and maximum values for NDVI and BT) can be estimated. Hence, agricultural drought indicators such as vegetation condition index (VCI), thermal condition index (TCI), and

vegetation health index (VHI) can be derived pixel-wise (*Kogan*, 1995b, 1990). The vegetation condition index is given by:

$$\text{VCI} = 100\frac{(\text{NDVI} - \text{NDVI}_{min})}{(\text{NDVI}_{max} - \text{NDVI}_{min})}, \quad \text{with} \quad \text{VCI} \in [0, 100]\,, \tag{2.21}$$

where NDVI is the weekly noise reduced NDVI, and NDVI_{min} and NDVI_{max} the multi-year weekly absolute minimum and maximum NDVI values, respectively. The thermal condition index is given by:

$$\text{TCI} = 100\frac{(\text{BT}_{max} - \text{BT})}{(\text{BT}_{max} - \text{BT}_{min})}, \quad \text{with} \quad \text{TCI} \in [0, 100]\,, \tag{2.22}$$

where BT is the weekly noise reduced BT, and BT_{min} and BT_{max} the multi-year weekly absolute minimum and maximum BT values, respectively. The vegetation health index is given by:

$$\text{VHI} = (\alpha)\text{VCI} + (1 - \alpha)\text{TCI}, \quad \text{with} \quad \text{VHI} \in [0, 100]\,, \tag{2.23}$$

where α is a weighting coefficient. While VCI is a proxy for the moisture condition and its lower values reflect a water-related stress, TCI is a proxy for the thermal condition and its lower values indicate a temperature and wetness-related stress. The composite index VHI is a linear combination of the former two indices to approximate the vegetation health (Fig. 2.3). VHI fluctuates annually between 0 (unfavourable condition) to 100 (favourable condition). The values of these indices above 100 and below 0 are clipped. Based on this definition of vegetation health, extreme agricultural drought can be defined when VHI < 26. Please note that vegetation stress detected by VHI could not be necessarily caused by a drought event, e.g., a change in the land cover can change the signal as well (*Sepulcre-Canto et al.*, 2012; *Van Loon et al.*, 2016b). Thus, VHI should be interpreted carefully.

For a summary on the validation and studies that use this dataset for agricultural droughts monitoring, we refer to *Yang et al.* (2020).

2.2.3.2 Preprocessing of Remote Sensing Data for Focal-TSMP in Chapter 5

In this section we describe the data preprocessing that is needed prior to apply DL in Chapter 5. Overall the TSMP has 30 years of data (1989-2019) (Section 2.2.2.1). We reserved the years 1989-2009 (AVHRR era) and 2013-2016 (VIIRS era) for training, 2010-2011 (AVHRR era) and 2017 (VIIRS era) for validation, and 2012 (AVHRR era) and 2018-2019 (VIIRS era) for testing. Due to the fact that the remote sensing data were obtained from two different satellite systems, the data derived from VIIRS have to be first adjusted to insure continuity and consistency with the data derived from AVHRR. *Yang et al.* (2018, 2021b) showed that the discrepancy between sensors are mainly due to the differences in spectral response ranges and calibration parameters. Compared to BT/TCI, this has a greater impact on NDVI/VCI (*Kogan et al.*, 2015). Considering this issue, we followed the same re-compositing approach as described in *Yang et al.* (2021b). The re-compositing approach can be used to generate cross-sensor vegetation products for the time period from 2013 to 2019. In fact NDVI/BT from different sensors can be decomposed into climatologies and VCI/TCI. The climatology provides information about the Ecosystem and it is sensor-specific. While VCI/TCI for the same ecosystem location are cross-sensor. Thus, using climatology from AVHRR and VCI from

VIIRS, Eq. (2.21) can be reformulated to re-composite NDVI for AVHRR as following:

$$\text{NDVI}'_{(AVHRR)} = \left(\frac{\text{VCI}_{(VIIRS)}}{100}\right)\left(\text{NDVI}_{(max,AVHRR)} - \text{NDVI}_{(min,AVHRR)}\right) + \text{NDVI}_{(min,AVHRR)}\,, \tag{2.24}$$

where $\text{NDVI}'_{(AVHRR)}$ is the converted weekly noise reduced NDVI from VIIRS to AVHRR, $\text{VCI}_{(VIRRS)}$ is the Vegetation Condition Index derived from VIIRS, $\text{NDVI}_{(min,AVHRR)}$ and $\text{NDVI}_{(max,AVHRR)}$ are the multi-year weekly absolute minimum and maximum NDVI values (climatology) derived from AVHRR, respectively. Similarly from Eq. (2.22) we have:

$$\text{BT}'_{(AVHRR)} = \text{BT}_{(max,AVHRR)} - \left(\frac{\text{TCI}_{(VIIRS)}}{100}\right)\left(\text{BT}_{(max,AVHRR)} - \text{BT}_{(min,AVHRR)}\right)\,, \tag{2.25}$$

where $\text{BT}'_{(AVHRR)}$ is the converted weekly noise reduced BT from VIIRS to AVHRR, $\text{TCI}_{(VIRRS)}$ is the Thermal Condition Index derived from VIIRS, $\text{BT}_{(min,AVHRR)}$ and $\text{BT}_{(max,AVHRR)}$ are the multi-year weekly absolute minimum and maximum BT values (climatology) derived from AVHRR, respectively. Please note that $\text{VCI}_{(VIIRS)}$ and $\text{TCI}_{(VIIRS)}$ were based on a pseudo long-term VIIRS climatology (for more details on this, see *Yang et al.* (2018)). In addition, the TSMP simulation and target remote sensing data have to be spatially aligned in the same domain. After the continuity at NDVI and BT level has been realized, we mapped these two products into the TSMP rotated coordinate system over the EURO-CORDEX EUR-11 domain (Section 2.2.2.1).

For the mapping, we up-scaled the data from 0.05° to 0.11° resolution based on a first-order conservative mapping (*Jones*, 1999) using the package from *Zhuang et al.* (2020). For calculating the spatial mean, we excluded invalid, water, and coastal lines pixels. Afterwards, we computed VCI, TCI and VHI based on Eq. (2.21)-(2.23). We note that the weighted coefficient α in Eq. (2.23) can be empirically calibrated as a spatially variant factor (*Zeng et al.*, 2022, 2023). Following previous works, we set α to its standard value 0.5 in all experiments as in *Yang et al.* (2020). Furthermore, masks over desert and very cold areas were extracted from the quality assurance (QA) metadata provided with the data. As mentioned in Section 2.2.3, a temporal decomposition was conducted to remove some discontinuity and aggregate the data into a weekly product. However, some pixels will still be empty. To tackle this issue, we first checked if the pixel was covered by another satellite and averaged the measurements of the satellites. If it was not the case, we flagged the pixel as invalid and discard it from the training and evaluation. Eventually, the preprocessed data are aggregated into data cubes ({variable, lat, lon}) on a weekly basis and stored as netCDF files. This remote sensing dataset can serve as a reference to train and evaluate the DL model performance. Overall, this includes 1263, 156, and 139 samples (weeks) for training, validation, and testing, respectively.

2.2.3.3 Preprocessing of Remote Sensing Data in Chapter 6

We follow the same preprocessing as in Section 2.2.3.2 to derive the agricultural drought indices (VCI, TCI, and VHI). The remote sensing dataset is provided on the Plate Carrée projection (geographic latitude and longitude). The target agricultural drought data and reanalysis data have to be aligned in the same coordinate systems and over the same regions. To realize this, we mapped the remote sensing data onto the Lambert conformal conical grid for CERRA (Section 2.2.5) and onto the rotated coordinate systems over the different CORDEX domains for ERA5-Land (Section 2.2.4). For the mapping, we use the first-order conservative mapping (*Jones*, 1999) using the software

from *Zhuang et al.* (2020). This remote sensing dataset serves as a reference of extreme agricultural drought events to train and evaluate the performance of the model in Chapter 6. The training, validation, and test splits of the remote sensing data follow the same ones shown in Table 2.7. Table 2.6 shows the ratio of extreme events in the datasets. Please note that there is no ground truth for drivers or anomalies in our real-world dataset (see Chapter 6). We only report the ratio of extreme agricultural drought events, which can be detected using remote sensing data.

Region	Domain	Extremes (%)	
		Val	Test
Europe	CERRA	4.34	5.32
Europe	EUR-11	3.20	2.86
Africa	AFR-11	6.41	6.87
North America	NAM-11	3.68	6.61
South America	SAM-11	5.16	6.53
Central Asia	CAS-11	3.60	4.38
East Asia	EAS-11	3.16	3.05

Table 2.6: Details regarding the ratio of extreme events in the pre-processed NOAA remote sensing data.

2.2.4 ERA5-Land Reanalysis Data

Reanalysis data aim to provide a coherent and complete reconstruction of the historical Earth system state as close to reality as possible. During reanalysis, short-term forecasts from numerical climate models are refined with observations within the so called data assimilation framework (*Hersbach et al.*, 2020). We conducted the experiments on a real-world ERA5-Land reanalysis dataset. The ERA5-Land reanalysis dataset is described in *Muñoz Sabater et al.* (2021) and the raw data were provided by the Copernicus Climate Change Service (C3S) Climate Data Store (CDS), `https://doi.org/10.24381/cds.e2161bac`. ERA5-Land is widely used for global climate research and it is provided hourly at $0.1^\circ \times 0.1^\circ$ on the regular latitude longitude grid.

2.2.4.1 Precprocessing of ERA5-Land in Chapter 6.

In Chapter 6, we aggregated the dataset on a weekly basis and selected the years within the period overlapping with the remote sensing data described in Section 2.2.3. In addition, we mapped ERA5-Land into 6 CORDEX domains (*Giorgi et al.*, 2009) over the globe and conduct experiments on each region separately.

We do the experiments with 6 common variables from ERA5-Land based on their connections to agricultural droughts. For all regions on ERA5-Land, we used the following variables: {"t2m", "fal", "e", "tp", "swvl1", "stl1"}. For each variable and week, we computed the mean and standard deviation separately. Technical details regarding the reanalysis dataset are provided in Tables 2.7, 2.8, and 2.9. CORDEX regions (*Giorgi et al.*, 2009) used in this study are shown in Fig. 2.4.

Dataset	Region	CORDEX	Resolution	Train	Val	Test
ERA5-Land	Europe	EUR-11	412×424	1981-2017	2018-2020	2021-2024
ERA5-Land	Africa	AFR-11	804×776	1981-2017	2018-2020	2021-2024
ERA5-Land	North America	NAM-11	520×620	1981-2017	2018-2020	2021-2024
ERA5-Land	South America	SAM-11	668×584	1981-2017	2018-2020	2021-2024
ERA5-Land	Central Asia	CAS-11	400×612	1981-2017	2018-2020	2021-2024
ERA5-Land	East Asia	EAS-11	668×812	1981-2017	2018-2020	2021-2024

Table 2.7: Datasets used in the experiments on real-world data. CORDEX domains are defined based on *Giorgi et al.* (2009).

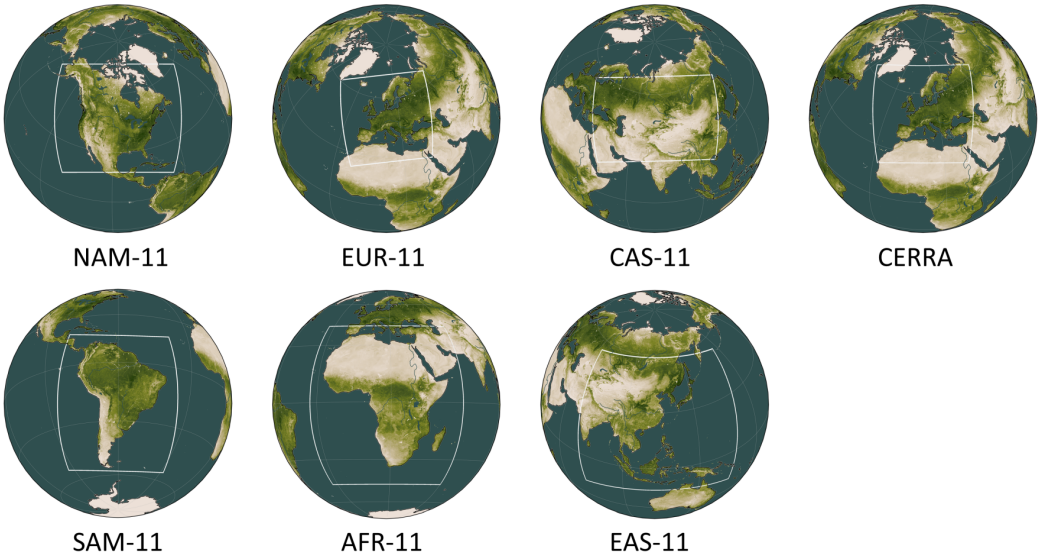


Figure 2.4: The definition of the domains used in the dissertation. ERA5-Land reanalysis is mapped onto the CORDEX domains (*Giorgi et al.*, 2009). CERRA reanalysis (Section 2.2.5) has its own domain definition (*Schimanke et al.*, 2021).

2.2.4.2 Precprocessing of ERA5-Land in Chapter 7.

In Chapter 7, in contrast to operational GloFAS which uses ERA5 as forcing data, we use ERA5-Land as an initial land surface condition for the forecast. We processed 14 instantaneous state variables at 00:00 UTC and 18 daily accumulated state variables (00:00 UTC previous day to 00:00 UTC current day). More details are provided in Table 2.8. As ERA5-Land is provided at $0.1^\circ \times 0.1^\circ$. We mapped the data onto the GloFAS regular latitude and longitude (Plate Carrée projection) using bilinear mapping and implemented by *Zhuang et al.* (2020).

2.2.5 CERRA Reanalysis Data

We conducted the experiments on real-world reanalysis CERRA dataset (*Schimanke et al.*, 2021). CERRA is a regional reanalysis for Europe and is provided originally at 5.5km × 5.5km on its Lambert conformal conical grid with a 3-hourly temporal resolution. We aggregated the dataset on a weekly basis and selected the years within the period overlapping with the remote sensing

Variable	Long name	Unit	Height	Surface parameters
d2m	2m dewpoint temperature	K	2m	instantaneous
e	total evaporation	m of water equivalent	surface	accumulated
es	snow evaporation	m of water equivalent	surface	accumulated
evabs	evaporation from bare soil	m of water equivalent	surface	accumulated
evaow	evaporation from open water surfaces excluding oceans	m of water equivalent	surface	accumulated
evatc	evaporation from the top of canopy	m of water equivalent	surface	accumulated
evavt	evaporation from vegetation transpiration	m of water equivalent	surface	accumulated
lai_hv	leaf area index high vegetation	m^2/m^2	2m	instantaneous
lai_lv	leaf area index low vegetation	m^2/m^2	2m	instantaneous
pev	potential evaporation	m	2m	accumulated
sf	snowfall	m of water equivalent	surface	accumulated
skt	skin temperature	K	surface	instantaneous
slhf	surface latent heat flux	J/m^2	surface	accumulated
smlt	snowmelt	m of water equivalent	surface	accumulated
sp	surface pressure	Pa	surface	instantaneous
src	skin reservoir content	m of water equivalent	surface	instantaneous
sro	surface runoff	m	surface	accumulated
sshf	surface sensible heat flux	J/m^2	surface	accumulated
ssr	surface net solar radiation	J/m^2	surface	accumulated
ssrd	surface solar radiation downwards	J/m^2	surface	accumulated
ssro	subsurface runoff	m	subsurface	accumulated
stl1	soil temperature	K	soil layer (0 - 7 cm)	instantaneous
str	surface net thermal radiation	J/m^2	surface	accumulated
strd	surface thermal radiation downwards	J/m^2	surface	accumulated
swvl1	volumetric soil water	m^3/m^3	soil layer (0 - 7 cm)	instantaneous
swvl2	volumetric soil water	m^3/m^3	soil layer (7 - 28 cm)	instantaneous
swvl3	volumetric soil water	m^3/m^3	soil layer (28 - 100 cm)	instantaneous
swvl4	volumetric soil water	m^3/m^3	soil layer (100 - 289 cm)	instantaneous
t2m	2m temperature	K	2m	instantaneous
tp	total precipitation	m	surface	accumulated
u10	10 metre U wind component	m/s	10m	instantaneous
v10	10 metre V wind component	m/s	10m	instantaneous

Table 2.8: Details about the processed variables from ERA5-Land reanalysis (*Muñoz Sabater et al.*, 2021) in Chapter 7.

Dataset	Variable name	Long name	Unit	Height
ERA5-Land	d2m	2m dewpoint temperature	K	2m
	t2m	2m temperature	K	2m
	fal	forecast albedo	%	surface
	skt	skin temperature	K	surface
	stl1	soil temperature	K	soil layer (0 - 7 cm)
	sp	surface pressure	Pa	surface
	e	total evaporation	m of water equivalent	above ground
	tp	total precipitation	m	surface
	swvl1	volumetric soil water	m^3/m^3	soil layer (0 - 7 cm)

Table 2.9: Details regarding the processed variables from ERA5-Land (*Muñoz Sabater et al.*, 2021) reanalysis in Chapter 6.

data described in Section 2.2.3. In Chapter 6, we do the experiments with 6 common variables from CERRA based on their connections to agricultural droughts. More specifically, we did the experiments with: {"t2m", "al", "tcc", "tp", "vsw", "r2"}. For each variable and week, we computed the mean and standard deviation separately. More details regarding the processed variables and the domains along with the training/validation/test splits are provided in Tables 2.10 and 2.11. The raw reanalysis data were provided by the Copernicus Climate Change Service (C3S) Climate Data Store (CDS), CERRA on single levels, `https://doi.org/10.24381/cds.622a565a`.

Dataset	Region	Resolution	Train	Val	Test
CERRA	Europe	1069×1069	1984-2015	2016-2018	2019-2021

Table 2.10: CERRA reanalysis used in the experiments on real-world data in Chapter 6.

2.2.6 GloFAS

The Global Flood Awareness System (GloFAS) is an operational system developed by the European Commission's Joint Research Centre (JRC) and operated by ECMWF under the Copernicus Emergency Management Service (CEMS) (*Alfieri et al.*, 2013). It provides real-time global-scale flood forecasts and a long-term hydrological reanalysis dataset, a key resource for flood risk assessment, climate impact studies, and machine learning applications. Fig. 2.5 shows the workflow of GloFAS to forecast river discharge and flood events. The GloFAS-ERA5 reanalysis is the long-term retrospective component of GloFAS (*Harrigan et al.*, 2020). It delivers daily river discharge estimates from 1979 to present at a spatial resolution of 0.05° ($\sim$ 5 km) and a global coverage (90°N-60°S, 180°W-180°E). The reanalysis is generated by coupling surface and subsurface runoff from the ERA5 reanalysis (*Hersbach et al.*, 2020), produced by the H-TESSEL and surface model (*Balsamo et al.*, 2009) with the LISFLOOD hydrological and river routing model (*Knijff et al.*, 2010). While ERA5 runoff is computed at $\sim$ 31 km resolution and lacks spatial connectivity, it is downscaled to 0.05° using a nearest-neighbour approach and routed through LISFLOOD to simulate realistic river discharge (dis24, in $m^3 s^{-1}$) across the global river network. The daily GloFAS reanalysis discharge data represents the mean value between 00:00 UTC previous day and 00:00 UTC current

Dataset	Variable name	Long name	Unit	Height
CERRA	al	albedo	%	surface
	hcc	high cloud cover	%	above 5000m
	lcc	low cloud cover	%	surface-2500m
	mcc	medium cloud cover	%	2500m-5000m
	liqvsm	liquid volumetric soil moisture	m^3/m^3	top layer of soil
	msl	mean sea level pressure	Pa	surface
	r2	2 metre relative humidity	%	2m
	si10	10 metre wind speed	m/s	10m
	skt	skin temperature	K	surface
	sot	soil temperature	K	top layer of soil
	sp	surface pressure	Pa	surface
	sr	surface roughness	m	surface
	t2m	2 metre temperature	K	2m
	tcc	total Cloud Cover	%	above ground
	tciwv	total column integrated water vapour	kg/m^2	surface
	tp	total Precipitation	kg/m^2	surface
	vsw	volumetric soil moisture	m^3/m^3	top layer of soil
	wdir10	10 metre wind direction	○	10m

Table 2.11: Details regarding the processed variables from CERRA (*Schimanke et al.*, 2021) reanalysis used in Chapter 6.

day. Similarly to GloFAS, there exists an early warning system for Europe (EFAS) with higher resolution (*Dottori et al.*, 2017). In this dissertation, GloFAS v4.0 is used for a global application. The dataset is publicly available on the Climate Data Store and Early Warning Data Store (EWDS), `https://doi.org/10.24381/cds.a4fdd6b9`. Table 2.12 explains the details of four variables we took from the GloFAS reanalysis dataset as the model inputs in Chapter 7. The GloFAS-ERA5 reanalysis supports the derivation of flood thresholds (e.g., 2-, 5-, and 20-year return periods) and serves as the initial condition for real-time forecasts such as GloFAS-30d and GloFAS-Seasonal. More details about GloFAS can be found in *Harrigan et al.* (2020).

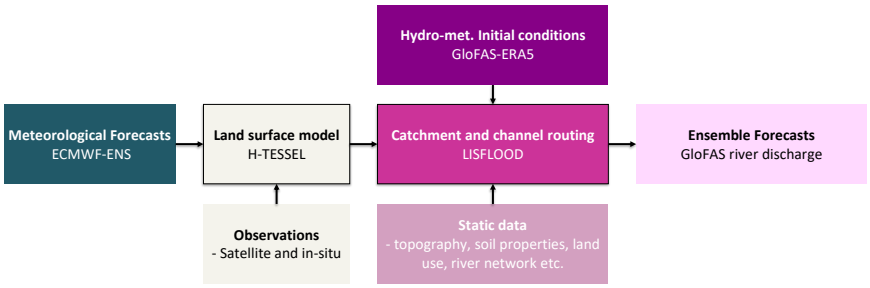


Figure 2.5: An overview of the key modules in the GloFAS forecasting system. GloFAS-ERA5 reanalysis uses ERA5 meteorological reanalysis data instead of ECMWF ensemble forecasts (ENS). Figure outline from *Harrigan et al.* (2020).

Despite its broad applicability, the GloFAS-ERA5 reanalysis is subject to several limitations that researchers should be aware of. Regional biases have been identified, which may stem from uncertainties in the meteorological forcing provided by ERA5, the representation of runoff generation processes within the H-TESSEL land surface model, and limitations in the calibration of the LISFLOOD routing model. When sufficient observational discharge data are available, the LISFLOOD model is calibrated locally for each river catchment larger than 500 km^2, and each calibrated catchment has its own optimized parameter set. This could give the model better performance at the local scale but reduce its generalization ability. Additionally, anthropogenic influences such as dams and reservoirs are incorporated using simplified operational rules, largely due to the lack of globally available real-time release data. Finally, as the dataset is entirely driven by ERA5, it inherits known deficiencies of the reanalysis, including biases in precipitation and the absence of river discharge data assimilation, which may affect the realism of simulated hydrological conditions in some regions.

Variable	Long name	Unit	Height	Surface parameters
acc_rod24	runoff water equivalent	kg/m^2	surface and subsurface	accumulated
dis24	river discharge in the last 24 hours	m^3/s	surface	averaged over 24 hours
sd	snow depth water equivalent	kg/m^2	surface	instantaneous
swi	soil wetness index	-	root zone	instantaneous

Table 2.12: Details about the processed variables from GloFAS reanalysis (*Harrigan et al.*, 2020).

2.2.7 GRDC Observational River Discharge Data

We obtain observational river discharge from the Global Runoff Data Centre (GRDC) which is an international data repository that provides access to quality-controlled river discharge observation data from around the world. The GRDC dataset contains time series of daily and monthly river discharge data from over 10000 hydrological gauging stations across more than 160 countries from small headwater catchments (~ 10 km^2 drainage area) to very large river catchment like the Amazon river (5 million km^2 drainage area). GRDC data can be obtained from `https://grdc.bafg.de/`. All GRDC daily time series measured the value set at 00:00 of the beginning of the day (left-labeled). We used the GRDC dataset as the benchmark to evaluate the model performance in Chapter 7 and followed a similar data processing workflow as in *Harrigan et al.* (2023) and *Nearing et al.* (2024). We first removed the catchments with a drainage area smaller than 500 km^2 and obtained 5524 GRDC stations to avoid very big discrepancies between the drainage area defined in GRDC and GloFAS dataset (2.2.6). Next, we geo-located the GRDC stations to compare them with the GloFAS drainage network and removed the GRDC stations with more than 10% of drainage area differences. For geo-location, we projected the points on the GloFAS grid, compared each point with its 9 nearest points, and took the location with the highest KGE value. Finally, the GRDC stations with no ERA5-Land reanalysis data were discarded. This resulted in 3366 stations for the global evaluation. This narrowed down the global median drainage area difference to 2.21% with an interquartile range of 0.86% to 4.73%. The discharge observations are recorded at a daily time scale, with the unit $m^3 s^{-1}$ and converted from local time zone to 00:00 UTC via linear interpolation. For evaluation, the GRDC observational time series, which are originally left-labeled, were explicitly converted to

right-labeled time series to ensure a temporal consistency with the right-labeled predictions from the GloFAS simulations, RiverMamba, LSTM baseline, and Google reforecast. The F1 scores reported for the GRDC evaluation are based on synchronized detection windows between model predictions and observations.

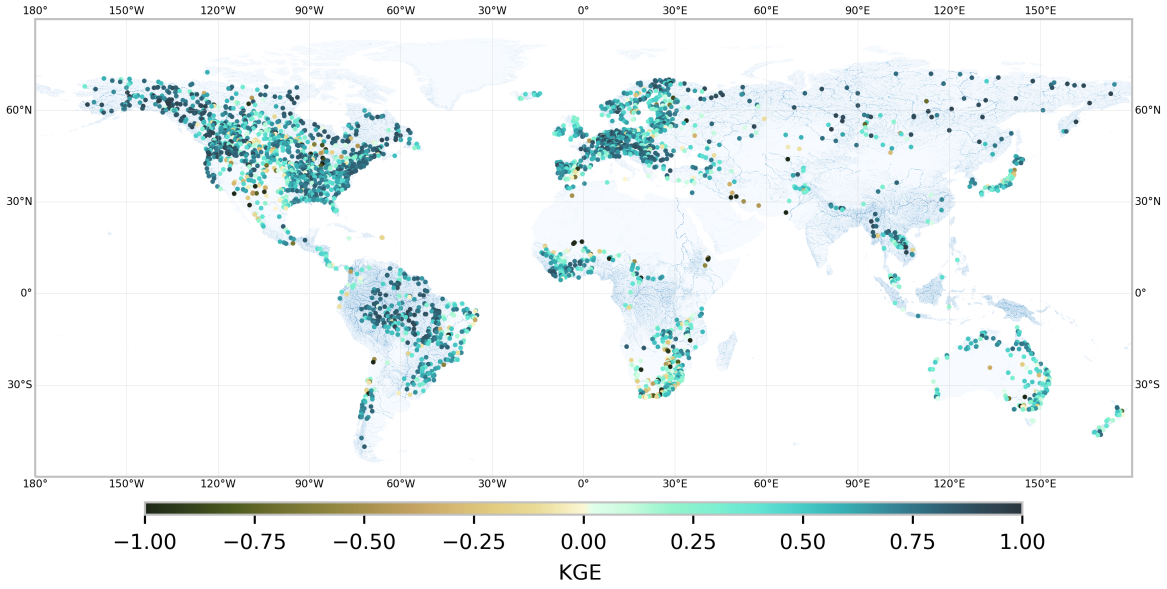


Figure 2.6: Locations of the selected 3366 GRDC stations used for training and evaluating the RiverMamba model. The colorbar shows the KGE value of GloFAS reanalysis discharge data against the GRDC discharge observations.

Fig. 2.6 shows the KGE values of GloFAS reanalysis data against the GRDC observation at the locations of all 3366 selected stations. In general, there is a good agreement between GloFAS and GRDC data globally, with a median KGE at 0.61 and an interquartile range of 0.36 to 0.77. In regions like south America, south Africa and Australia, the GloFAS reanalysis data have more inconsistency compared to the observations. Fig. 2.7 shows an exemplar hydrograph for a gauged station.

It is important to note that, compared to the GRDC observations, the GloFAS reanalysis dataset only simulates the naturalized flow without considering realistic human interventions such as dams, reservoirs, diversions, irrigation withdrawals, and other water management practices, and this can be a major source of bias in GloFAS compared to the GRDC data.

2.2.8 HRES Meteorological Forecast

Meteorological conditions serve as the driving forces behind hydrological processes. These are necessary for forecasting the river discharge and potential floods. We use the deterministic forecast of the ECMWF Integrated Forecast System (IFS) High Resolution (HRES) atmospheric model. The HRES data were obtained from the ECMWF Archive Catalogue, `https://www.ecmwf.int/en/forecasts/dataset/operational-archive`. We use HRES up to 7 days lead time and once per day at 00:00 UTC. The processed data are similar to *Nearing et al.* (2024) except that we used total evaporation as an additional forcing variable. In addition, we do not use any forecast for nowcasting at time step t. The processed data include 2 instantaneous and 5 daily accumulated vari-

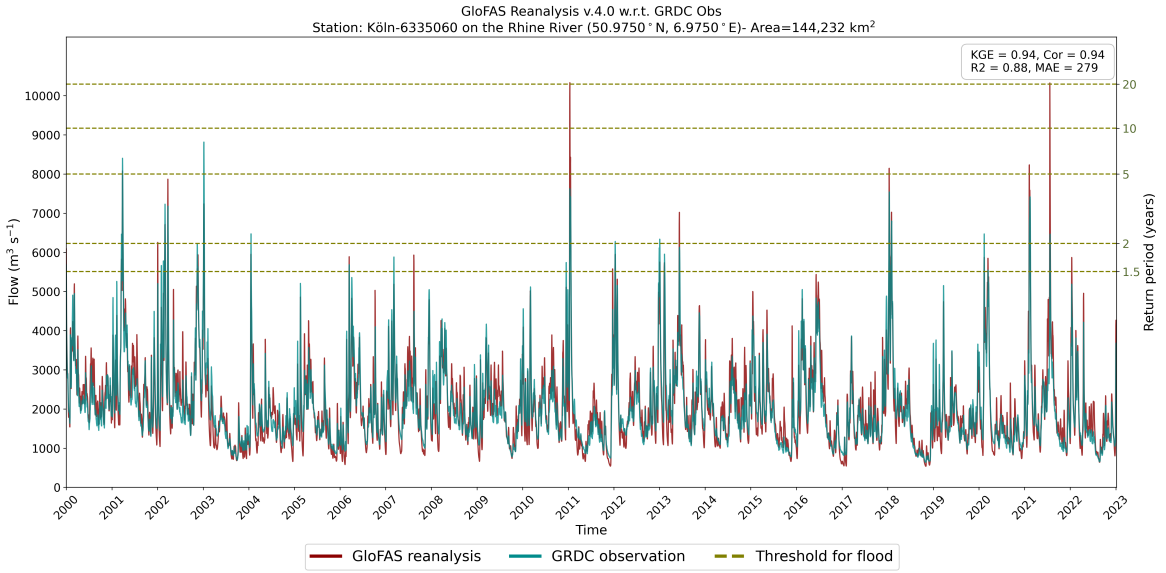


Figure 2.7: Hydrograph for GloFAS-reanalysis (red line) from 1 January 2000 to 31 December 2023 and observations (dark cyan line), for a gauging station on the Rhine River. The top-right box displays summary statistics from the reanalysis's evaluation against the observations.

ables at the surface level forecasts. Technical details regarding the meteorological forcing variables are provided in Tables 2.13.

While the operational archive provides data from 1985, the quality and the resolution of the forecasts in the earlier years are not sufficient for the application in Chapter 7. Therefore, we only processed and used data from 2010 to 2024. Data before 2010 were replaced by ERA5-Land. To match the resolution of the target GloFAS grid, we regridded HRES to $0.05^\circ \times 0.05^\circ$ regular grid.

Variable	Long name	Unit	Height	Surface parameters
e	total evaporation	m of water equivalent	surface	accumulated
sf	snowfall	m of water equivalent	surface	accumulated
sp	surface pressure	Pa	surface	instantaneous
ssr	surface net solar radiation	J/m^2	surface	accumulated
str	surface net thermal radiation	J/m^2	surface	accumulated
t2m	2m temperature	K	2m	instantaneous
tp	total precipitation	m	surface	accumulated

Table 2.13: Details about the processed variables from the ECMWF Integrated Forecast System (IFS) High Resolution (HRES) atmospheric model.

2.2.9 CPC Precipitation Data

Relying solely on the precipitation products from ERA5-Land reanalysis (Section 2.2.4) makes the model prune to the biases of the data assimilation which was used to derive the reanalysis. We use precipitation estimates as observational input from the National Oceanic and Atmospheric Administration (NOAA), Climate Prediction Center (CPC). The product is called Global Unified

Gauge-Based Analysis of Daily Precipitation. The CPC precipitation product is accumulated daily and provided globally at $0.5^\circ \times 0.5^\circ$. To match the resolution of the river discharge (Section 2.2.6), we mapped CPC data onto the GloFAS domain using nearest point algorithm which preserves the original coarse grid structure but refines the resolution. We did not do any modification for the CPC time zones. More details regarding the construction of the daily gauge analysis, the interpolation algorithm, and the gauge algorithm evaluation can be found in *Xie et al.* (2010), *Xie et al.* (2007), and *Chen et al.* (2008). Operational CPC data can be obtained from `https://psl.noaa.gov/data/gridded/data.cpc.globalprecip.html`.

2.2.10 LISFLOOD Static Data

River attributes and static maps are crucial to capture the sub-grid variability for the river discharge. For consistency and to make a fair comparison with GloFAS, we used LISFLOOD input static maps (*Choulga et al.*, 2024) similar to the operational GloFAS. This includes 96 time-invariant variables from 7 different categories (Table 2.14). The maps are provided at the same resolution as GloFAS at 3 arcmin and covering the globe (90°N-60°S, 180°W-180°E). We excluded the lakes, reservoirs and some static water demand maps.

The LISFLOOD static maps can be obtained from the Joint Research Centre Data Catalogue at `http://data.europa.eu/89h/68050d73-9c06-499c-a441-dc5053cb0c86`.

Category	# Static variables
catchment morphology and river network	12
grid	2
land use	6
vegetation properties	45
soil properties	14
water demand	3
GloFASv4.0 calibrated parameters	14

Table 2.14: The processed LISFLOOD static and parameter maps (*Choulga et al.*, 2024).

2.2.11 HydroRIVERS Static Data

HydroRIVERS data (*Lehner and Grill*, 2013; *Linke et al.*, 2019) are widely used to train deep learning models in hydrology. To run experiments on HydroRIVERS, we obtained static river attributes from `https://www.hydrosheds.org/products/hydrorivers`. The data is stored as a shape file. To map them onto the GloFAS domain, we first extract the coordinates of the rivers and then project them with the grid points on the WGS-84 ellipsoid (Eq. 7.1 and 7.2). Then, for each GloFAS grid point, depending on the attribute type, we either average the attributes or take the most frequent attribute within a radius of 5 km. If no attributes were found, we increase the radius to 12 km, and 24 km, respectively. We processed 299 river feature attributes overall and experimented with 103 features, e.g., we removed the monthly attribute statistics from the static features. Fig. 2.8 gives an overview of the processed HydroRIVERS data.

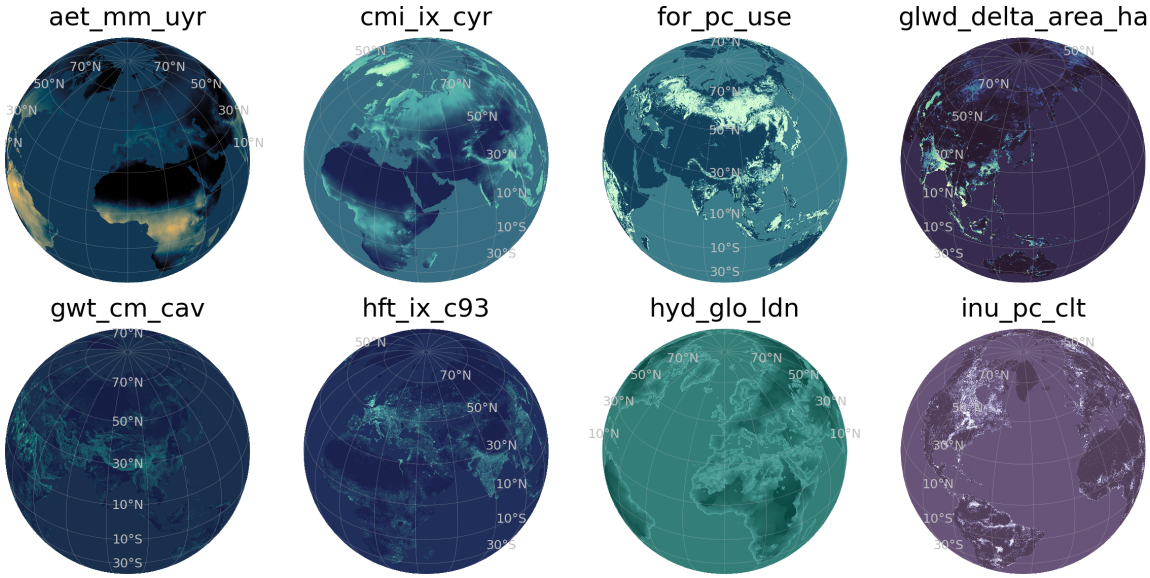


Figure 2.8: An overview of the processed HydroRIVERS static features. HydroRIVERS is mapped into the GloFAS domain (global regular latitude-longitude grid).

2.2.12 Diagnostic River Points for Flood Forecasting

The original resolution of GloFAS v.4.0 (Section 2.2.6) is 3 arcmin with an image resolution of 3000×7200 (21 million pix). In order to run experiments efficiently in Chapter 7, we sampled points. For this, we remove all points that are not located on the land surface, e.g., points over ocean or sea. This reduced the points from $21,000,000$ to $6,221,926$ points. We excluded points with median river discharge less than $10\ m^3 s^{-1}$ since river discharge is more relevant where there is a water flow, e.g., points that are located near to water bodies and not located over desert or glacier regions. Points which are close to rivers (distance 1 pix to points with discharge > $10\ m^3 s^{-1}$) were not excluded. We also do not exclude points defined as GRDC stations. This reduced the points further to $1,529,667$ diagnostic river points on which we train and test. Figure 2.9 gives an overview of the filtered diagnostic river points used in Chapter 7.

2.3 Return Periods and Flood Definition

In hydrology, the return period (also known as the recurrence interval) is a statistical measure that estimates how often a given hydrological event such as a flood, drought, or heavy rainfall is expected to occur on average over a long period. In this study, the return periods refer to the flood frequency. For example, a 2-year flood has a 50% chance of being exceeded in any given year. The return period (RP) is defined as the inverse of the annual exceedance probability (AEP):

$$RP = \frac{1}{AEP}\,. \tag{2.26}$$

In practice, flood return periods are used to define flood thresholds, e.g., a flood warning is triggered when discharge exceeds the 2-year threshold. Note that a high return period, e.g., 20 years does not necessarily imply actual flooding in all regions, particularly in highly regulated or

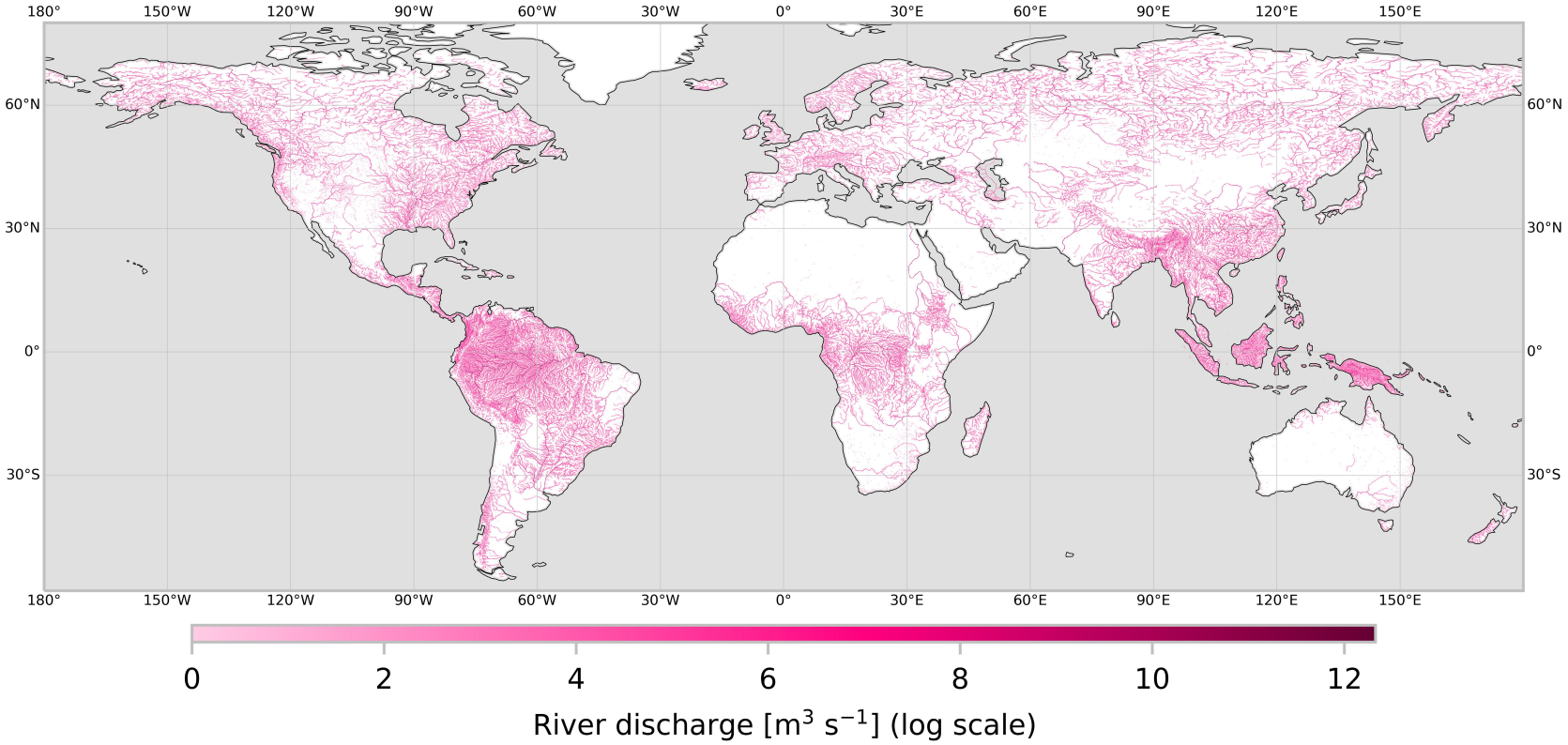


Figure 2.9: Overview of the selected river points.

flood-resilient areas. A high return period event simply reflects the statistical rarity in streamflow magnitude, and should not be equated with a flood event without additional context (e.g., thresholds, inundation). The return period is used here as a proxy indicator of hydrological extremity, which we call flood severity. In this study, we adapted the GloFAS approach to define flood thresholds corresponding to selected return periods (or recurrence intervals) of 1.5, 2, 5, 10, 20, 50, 100, 200, and 500 years. These thresholds are derived from the LISFLOOD reanalysis simulations, which are forced with ERA5 meteorological data. The return levels are estimated by fitting a Gumbel extreme value distribution to the annual maxima for the period 1979-2022, using the L-moments method. For the evaluation on GloFAS reanalysis data, we use the pre-defined return periods data from the Copernicus Emergency Management Service, https://confluence.ecmwf.int/display/CEMS/Auxiliary+Data. Fig. 2.7 shows the flood thresholds defined by different return periods. For the GRDC observation dataset, we calculated the return periods at individual stations from the first available observation date to 2022. To allow a fair evaluation of GloFAS reanalysis data on GRDC observation, the return period of GloFAS data is also calculated at GRDC stations but on the local available observation time period. Note that while we calculated return period thresholds separately from both the GRDC observations and the GloFAS reanalysis, we did not calculate return periods from the trained ML models reforecast as this would require generating a long reforecast climatology to fit a statistical extreme value distribution.

2.4 PRUDENCE Scientific Regions

The PRUDENCE regions refer to the project (Prediction of Regional scenarios and Uncertainties for Defining EuropeaN Climate change risks and Effects). The project focuses on studies on high-resolution regional climate and climate change scenarios. The PRUDENCE regions contain 8 regions, Alps (AL), British Isles (BI), Eastern Europe (EA), France (FR), Iberian Peninsula (IP), Mediterranean (MD), Mid-Europe (ME), and Scandinavia (SC). We use these regions for the experiments in Chapter 5.

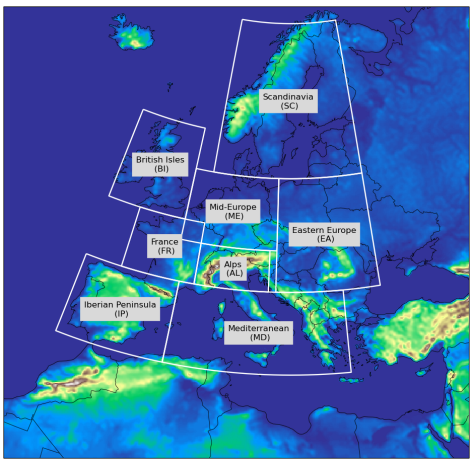


Figure 2.10: Orography over the EURO-CORDEX domain. The white boundaries with the labeled names inside define the PRUDENCE regions. In Chapter 5, we compute the time series for validating and testing agricultural drought indices over these regions.

2.5 Evaluation Metrics

In this section, we provide details about the evaluation metrics used to assess the performance of different methods in this dissertation.

We denote the ground truth value as $\hat{Y}$ and the prediction as Y. The overall mean value of the ground truth in a sample is denoted as $\hat{\bar{Y}}$ and the overall mean value of the prediction in a sample is denoted as $\bar{Y}$. We also denote P as the total number of pixels or points in the evaluated sample. In case of images as in Chapters 4, 6 and 7, $P = W \times H$, where H and W are the height and width of the image respectively.

2.5.1 Mean Absolute Error (MAE)

This metric represents the average of the absolute differences between the predicted and observed values. It provides a straightforward measure of model accuracy. MAE is less sensitive to outliers than RMSE:

$$\mathrm{MAE}(\hat{\mathbf{Y}}_p, \mathbf{Y}_p) = \frac{1}{P} \sum_{p=1}^{P} \left| \hat{\mathbf{Y}}_p - \mathbf{Y}_p \right| . \tag{2.27}$$

2.5.2 Root Mean Square Error (RMSE)

This metric measures the square root of the average squared differences between predicted and observed values. It penalizes large errors more heavily than MAE:

$$\mathrm{RMSE}(\hat{\mathbf{Y}}_p, \mathbf{Y}_p) = \sqrt{\frac{1}{P} \sum_{p=1}^{P} (\hat{\mathbf{Y}}_p - \mathbf{Y}_p)^2} . \tag{2.28}$$

2.5.3 Bias

We compute the Bias as:

$$\mathrm{Bias}(\hat{\mathbf{Y}}_p, \mathbf{Y}_p) = \mathbf{Y}_p - \hat{\mathbf{Y}}_p . \tag{2.29}$$

2.5.4 Pearson Correlation Coefficient (R_p)

Pearson correlation (R_p) is a parametric correlation that measures the linear correlation between the predicted and observed values:

$$R_p(\hat{\mathbf{Y}}_p, \mathbf{Y}_p) = \frac{\sum_{p=1}^{P} (\hat{\mathbf{Y}}_p - \hat{\bar{Y}})(\mathbf{Y}_p - \bar{Y})}{\sqrt{\sum_{p=1}^{P} (\hat{\mathbf{Y}}_p - \hat{\bar{Y}})^2} \sqrt{\sum_{p=1}^{P} (\mathbf{Y}_p - \bar{Y})^2}} . \tag{2.30}$$

The best value for R_p is 1 which represents a perfect positive correlation and the lowest value is -1 which represents a perfect negative correlation.

2.5.5 Spearman Correlation Coefficient (R_s)

Spearman correlation (R_s) is a non-parametric measure of relationship between predicted and observed values that can be calculated as follows:

$$R_s(\hat{\mathbf{Y}}_p, \mathbf{Y}_p) = R_p(\text{Rank}(\mathbf{Y}_p), \text{Rank}(\hat{\mathbf{Y}}_p))\,, \tag{2.31}$$

where Rank(Y_p) and Rank($\hat{Y}_p$) are ranks obtained from the predicted and observed values, respectively. A perfect positive correlation occurs when R_s is 1.

2.5.6 Coefficient of Determination (R2)

This coefficient metric measures the predictive power of a model relative to the observed mean and it is calculated as follows:

$$\text{R2}(\hat{\mathbf{Y}}_p, \mathbf{Y}_p) = 1 - \frac{\sum_{p=1}^{P}(\hat{\mathbf{Y}}_p - \mathbf{Y}_p)^2}{\sum_{p=1}^{P}(\hat{\mathbf{Y}}_p - \hat{\bar{Y}})^2}\,. \tag{2.32}$$

The highest value for R2 is 1 which represents a perfect fit, while values below 0 suggest that the model performs worse than using the observed mean. Please note that R2 measures the variability in $\hat{Y}_p$ predicted by the model thus it is by definition inversely proportional to the variance and noise in the observations and should be interpreted carefully. R2 has also the same meaning as Nash-Sutcliffe Efficiency (NSE) which is commonly used in hydrology.

2.5.7 Kling-Gupta Efficiency (KGE)

KGE is a composite metric that combines correlation, bias, and variability. It addresses some weaknesses of NSE by ensuring balance across multiple aspects of model performance. Like NSE, values near 1 indicate good performance, while values below 0 indicate performance worse than the observed mean:

$$\text{KGE}(\hat{\mathbf{Y}}_p, \mathbf{Y}_p) = 1 - \sqrt{(R_p - 1)^2 + (\text{BR} - 1)^2 + (\text{VR} - 1)^2}\,, \tag{2.33}$$

$$\text{BR} = \frac{\bar{Y}}{\hat{\bar{Y}}}, \quad \text{VR} = \frac{\text{CV}}{\hat{\text{CV}}}\,, \tag{2.34}$$

BR is the bias ratio, VR is the variability ratio, and $\hat{\text{CV}} = \hat{\sigma}/\hat{\bar{Y}}$ and $\text{CV} = \sigma/\bar{Y}$ are the coefficients of variation, where $\hat{\sigma}$ and σ are the standard deviations of the target and predicted values, respectively. The best value for KGE is 1.

2.5.8 Precision

Precision is the proportion of correctly identified positive cases (e.g., flood events) among all predicted positives. High precision indicates a low false-positive rate:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}\,, \tag{2.35}$$

where TP is the number of true positives and FP is the number of false positives.

2.5.9 Recall

Recall is the proportion of correctly identified positives among all actual positives. High recall indicates a low false-negative rate:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{2.36}$$

where FN is the number of false negatives.

2.5.10 F1-score

This metric is the harmonic mean of precision and recall, particularly useful in imbalanced classification tasks (e.g., flood detection where flood events are rare):

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{2.37}$$

2.5.11 Intersection over Union (IoU)

This metric is used to evaluate localization tasks. IoU is computed as the area of overlap between the predicted and target segments over the union area of the predicted and target segments:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{\text{TP}}{\text{TP+FP+FN}}. \tag{2.38}$$

2.5.12 Overall Accuracy (OA)

OA is the accuracy obtained on all correct predictions for all classes:

$$\text{OA} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{2.39}$$

where TN is the number of true negatives.

2.5.13 Area Under the Receiver Operating Characteristic Curve (AUROC)

This metric is used in binary classification tasks to evaluate how well the classifier can distinguish between the positive class and negative one. The AUROC represents the area under the Operating Characteristic Curve (ROC) which visualize the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings. Each point on the ROC represents the classifier performance at a specific threshold. AUROC summarizes the overall classifier performance within multiple thresholds in one value. An AUROC of 0.5 indicates a random guess classifier and an AUROC of 1.0 represents a perfect classifier.

CHAPTER 3

Related Works

In this chapter, we discuss relevant literature related to this dissertation. We first discuss approaches for extreme and anomaly detection from climate data. Then, we provide an overview about related works on forecasting extreme weather and climate events. Finally, we discuss some of the previous works on conditional normalization, multi-branch neural networks, temporal positional encoding, and state space models.

Contents

3.1 Anomaly Detection 43
3.1.1 Anomalies and Extremes Detection in Climate Data 43
3.1.2 Anomaly Detection Algorithms 44
3.2 Extreme Events Forecasting 45
3.2.1 Wildfire Danger Forecasting 45
3.2.2 Vegetation Health and Agricultural Drought Forecasting 46
3.2.3 River Discharge and Flood Forecasting 47
3.3 Techniques in Neural Networks 48
3.3.1 Conditional Normalization 48
3.3.2 Multi-Branch Neural Networks 49
3.3.3 Temporal Positional Encoding 50
3.3.4 State Space Model (SSM) and Mamba 50

3.1 Anomaly Detection

3.1.1 Anomalies and Extremes Detection in Climate Data

The identification of climatic changes and extreme weather has been a subject of many studies (*Das and Parthasarathy*, 2009; *Salcedo-Sanz et al.*, 2024). Typical algorithms for extreme events detection are built upon domain knowledge in setting usually empirical thresholds for the physical variables through sensitivity experiments (*Huang et al.*, 2015; *Liu et al.*, 2016b). Many works applied multivariate and statistical methods to detect extreme events such as droughts (*Huang et al.*, 2015; *Cao et al.*, 2022; *Sepulcre-Canto et al.*, 2012; *Zhang et al.*, 2021b, 2017). However, individual events are difficult to generalize across multiple events (*Sillmann et al.*, 2017) and predefined indicators become less effective with changing climate (*Yihdego et al.*, 2019). Thus, machine and deep learning methods have been proposed as an alternative to classical methods, e.g., for supervised anomaly detection (*Dahoui*, 2023; *Düben et al.*, 2021) and for the detection of extremes in climate data (*Liu et al.*, 2016b; *Flach et al.*, 2017; *Racah et al.*, 2017; *Prabhat et al.*, 2021; *Weirich-Benet et al.*, 2023).

3.1.2 Anomaly Detection Algorithms

In Chapter 6, we address the problem of identifying spatio-temporal relations between the impacts of extreme events and their drivers in climate and weather data. Examples of extreme events are extreme droughts, floods, or heatwaves. These events can be represented by their impacts on state variables. For example, agricultural drought can be measured by extremely low values of the vegetation health index (VHI) and floods can be detected by high values of river discharge. We consider anomalies in atmospheric/hydrological state variables (e.g. temperature or soil moisture) or land-atmosphere fluxes (e.g. evaporation) as drivers for such measurable impacts of extreme events. Since we focus on anomalies in land and meteorological data as drivers for extremes, we give an overview of approaches for anomaly detection and discuss their applicability to the task in Chapter 6.

One-class: The main stream in one-class anomaly detection is to model the distribution of the normal data during training and consider the deviation from the learned features as anomalies. This includes distance-based (*Schölkopf et al.*, 1999; *Liznerski et al.*, 2021; *Yi and Yoon*, 2020), patch-based (*Yi and Yoon*, 2020; *Wang et al.*, 2021b; *Defard et al.*, 2021), student-teacher (*Deng and Li*, 2022; *Salehi et al.*, 2021; *Bergmann et al.*, 2020; *Wang et al.*, 2021a), and embedding-based approaches (*Defard et al.*, 2021; *Reiss et al.*, 2021; *Rippel et al.*, 2021; *Roth et al.*, 2022; *Liu et al.*, 2023b). In general, the alignment between the anomaly type and the assumptions of the methods is the critical factor for their performance (*Han et al.*, 2022). One of the limiting factors to apply these methods to climate data is that they assume priori knowledge about what is considered as normal. Furthermore, not all detected anomalies are drivers of an extreme event.

Reconstruction-based: These methods assume that a trained model to reconstruct normal data will be unsuccessful in reconstructing anomalies, while it will reconstruct normal data well. Despite being widely applied for anomaly detection problems (*Strothmann et al.*, 2019; *Mishra et al.*, 2021; *Lee and Kang*, 2022; *Miranda et al.*, 2022; *Sabokrou et al.*, 2019; *Venkataramanan et al.*, 2020; *Astrid et al.*, 2021; *Liu et al.*, 2023a; *Yang et al.*, 2023; *Madan et al.*, 2024; *Wyatt et al.*, 2022; *He et al.*, 2024), these methods face the same problem as the one-class methods. In addition, many studies showed that anomalies can still be reconstructed by the trained model (*You et al.*, 2022).

Self-supervised learning: These methods rely on the hypothesis that a model trained for a pretext task on normal data will be successful only on similar normal data during inference (*Huang et al.*, 2023; *Luo et al.*, 2020; *Schlüter et al.*, 2022; *Wang et al.*, 2022b). In addition to the above discussed limitations, finding a suitable pretext task for anomaly detection is challenging. For instance, common tasks such as solving a jigsaw puzzle (*Wang et al.*, 2022b) will fail in homogeneous regions and it is hard to scale with climate variables.

Pseudo-anomaly: The intuition in pseudo-based anomaly detection is to convert the problem of unsupervised learning into a supervised one by synthesizing abnormal data during training (*Zavrtanik et al.*, 2021; *Collin and De Vleeschouwer*, 2021; *Li et al.*, 2021; *Pourreza et al.*, 2021; *Zaheer et al.*, 2020a; *Thakare et al.*, 2023; *Zhang et al.*, 2024c). Since these methods depend partially on the degree to which the proxy anomalies correspond to the unknown true anomalies (*Deng and Li*, 2022), applying these approaches to the task in Chapter 6 would require some knowledge about the coupling between the variables and extremes.

Multivariate anomaly detection: Multivariate approaches detect anomalies simultaneously in multiple data streams (*Xu et al.*, 2022; *Lee et al.*, 2021; *Zhong et al.*, 2025; *Tuli et al.*, 2022; *Zheng et al.*, 2024). The main difference to our task in Chapter 6 is that we aim to detect drivers across

multiple data streams that do not necessary occur simultaneously.

Multiple instance learning: Multiple instance learning (MIL) has been proposed for weakly supervised anomaly detection (*Zhong et al.*, 2019; *Zaheer et al.*, 2020b; *Lv et al.*, 2021a; *Feng et al.*, 2021; *Park et al.*, 2023; *Zhou et al.*, 2023; *Qi et al.*, 2022; *Wu et al.*, 2022; *Sapkota and Yu*, 2022). In MIL-based algorithms, the model is provided with labeled positive and negative bags where each bag includes a set of instances. The model is then trained to classify the instances inside the bags giving only the high level supervision, e.g., label of the bag (*Jiang et al.*, 2023). A weakly supervised approach has been also applied for hyperspectral anomaly detection (*Jiang et al.*, 2022b). Most algorithms such as (*Sultani et al.*, 2018; *Wan et al.*, 2020; *Tian et al.*, 2021; *Chen et al.*, 2023) choose the top-k potentially anomalous snippets within each video. This makes it challenging to apply them since the abnormality ratio varies in real-world applications (*Zhou et al.*, 2024). Furthermore, the MIL detector can be biased toward a specific class depending on the context (*Lv et al.*, 2023).

3.2 Extreme Events Forecasting

In what follows, we briefly review related works on forecasting extreme weather and climate events.

3.2.1 Wildfire Danger Forecasting

Wildfire forecasting or wildfire-susceptibility mapping from remote sensing and Earth observations data is a very important topic for wildfire management (*Jain et al.*, 2020). We briefly review some prior related works in this direction.

Typical methods include calculating an index like the Fire Weather Index (FWI) (*Van Wagner et al.*, 1985). Such index is a meteorologically based one meaning that it aims to estimate the fire danger by taking into the account the effects of relative humidity, wind speed, temperature and other weather variables on the occurrence and behavior of fires. The Copernicus Emergency Management Service (CEMS) together with ECMWF operates the Global Fire Forecasting system (GEFF) (*Giuseppe et al.*, 2016; *Di Giuseppe et al.*, 2024) which is a widely utilized system. In another line of work, *Iban and Sekertekin* (2022), *Pham et al.* (2020), *Gholami et al.* (2020), and *Wu et al.* (2023) relied on traditional machine learning approaches to generate susceptibility maps. *Shang et al.* (2020) and *Mitsopoulos and Mallinis* (2017) utilized Random Forests classifiers (RF). In their works, they studied the importance of biotic and abiotic predictors for wildfire forecasting. *Jiang et al.* (2021) proposed a deep learning approach based on a Multi-Layer Perceptron (MLP) and included a comparison with traditional machine learning algorithms. In *Le et al.* (2021), a similar MLP-based approach was presented to generate a forest fire danger map. *Zhang et al.* (2019) used a convolutional neural network (CNN) and extended their work later to predict fire susceptibility at the global level (*Zhang et al.*, 2021a). Other works with CNNs were conducted in *Bjånes et al.* (2021) and *Bergado et al.* (2021). Furthermore, *Huot et al.* (2022) approached the problem as a scene classification task using U-Net models to predict wildfire spreading. Their approach operates directly on the whole scene. A similar approach based on a U-Net++ model for global wildfire forecasting was proposed in *Prapas et al.* (2022a). *Yoon and Voulgaris* (2022) presented an approach that relies on a recurrent network with Gated Recurrent Units (GRU) to model past observations and on a CNN to predict wildfire probability maps for multiple time steps. Recently, *Prapas et al.* (2021) and *Kondylatos et al.* (2022) proposed to use LSTM-based (Long-Short Term Memory) approaches. They exploited both tem-

poral and spatio-temporal context by applying recurrent LSTM and ConvLSTM models. They did not consider the whole scene at once but rather the classification of one pixel at a time (pixel-level). Other works have proposed GNNs for wildfire forecasting. *Zhao et al.* (2024) proposed a causal graph and forecast wildfire risk eight days in advance in the Mediterranean. In *Michail et al.* (2025), GraphCast (*Lam et al.*, 2022) which is a graph-based neural network for weather forecast, has been used for seasonal wildfire forecasting. Transformers with self-attention were also utilized for global-scale wildfire forecasting. For example, *Prapas et al.* (2023) and *Zhu et al.* (2025) investigated the complex links between global climate patterns and wildfire events with self-attention.

Unlike these works, in Chapter 4, we do not treat all observation variables in the same way, but we propose a deep learning model that handles different types of variables in separated 2D and 3D CNN branches. In contrast to a ConvLSTM (*Prapas et al.*, 2021; *Kondylatos et al.*, 2022), which models spatial and temporal relations separately, a 3D CNN models spatio-temporal relations jointly. We assume that the dataset contains static and dynamic variables, which we argue is the case for most datasets. Similar to *Prapas et al.* (2021) and *Kondylatos et al.* (2022), we also formulate the problem as pixel-level classification taking into account the spatio-temporal local context around the target pixel.

3.2.2 Vegetation Health and Agricultural Drought Forecasting

A plenitude of studies exist about vegetation health prediction and forecasting from Earth observations. Unlike hydro-meteorological variables that can be predicted or forecast using numerical weather models, vegetation products demand an extended modeling representation of the surface and sub-surface (*Lees et al.*, 2022). There exist climate-vegetation models which directly simulate the vegetation dynamic based on ecological processes and statistical modeling. Nevertheless, they are limited by the complexity of the processes and poor generalization (*Chen et al.*, 2021). Recently, *Salakpi et al.* (2022a,b) predicted short-term vegetation products based on previous vegetation conditions and observational anomaly indices in a Bayesian auto-regressive approach. However, the interaction between vegetation and climate variability exhibits a strong non-linear behavior. Recently, DL models have become popular to build a predictive model for tasks that include complex or intractable cause and effect relations within the Earth system (*Bergen et al.*, 2019; *Tuia et al.*, 2024; *de Burgh-Day and Leeuwenburg*, 2023). In this respect, many studies explored the applicability of DL for vegetation health prediction using climate models and remote sensing data (*Das and Ghosh*, 2016; *Adede et al.*, 2019; *Ferchichi et al.*, 2022; *Wu et al.*, 2020; *Kraft et al.*, 2019; *Prodhan et al.*, 2021). A common approach is to use past vegetation conditions to predict the short-term future variations (*Nay et al.*, 2018; *Yu et al.*, 2022b; *Hammad and Falchetta*, 2022; *Lees et al.*, 2022; *Vo et al.*, 2023). In a related work, *Requena-Mesa et al.* (2021) addressed the problem of optical satellite imagery forecasting as a guided video prediction task. In their framework, vegetation dynamics approximated by NDVI is modeled at high resolution using past satellite images as initial conditions and static and reanalysis data as a model guidance. Similar approaches with this framework were presented in (*Robin et al.*, 2022; *Kladny et al.*, 2022; *Diaconu et al.*, 2022) and on a continental scale in (*Benson et al.*, 2024). While these works differ in their methodologies, e.g., in the predicted vegetation products, model architectures, and spatio-temporal resolutions, they have overall a good performance for short-term forecasting. Although short-term forecasting, e.g., for a few weeks, is very useful for short-term planning, a more significant contribution could be achieved with a much

longer forecasting time (*Marj and Meijerink*, 2011). Nonetheless, only few studies addressed long-term vegetation conditions forecasting (*Marj and Meijerink*, 2011; *Miao et al.*, 2015; *Patil et al.*, 2017; *Chen et al.*, 2021; *Wei et al.*, 2023). In addition, most studies focused only on a single indicator. The combination of different indicators like NDVI and BT with their corresponding drought indices provides complementary information on the vegetation state and is beneficial for vegetation monitoring (*Yang et al.*, 2020).

In Chapter 5, we propose a Focal-TSMP, a deep learning model that can forecast multiple vegetation products from a regional climate simulation at a continental scale and with a long-term forecast.

3.2.3 River Discharge and Flood Forecasting

3.2.3.1 Flood Forecasting

Floods can be categorized into three common types. The first type is the *fluvial* or *riverine flood* (*Bomers and Hulscher*, 2023). It occurs when the water level in a stream rises and overflows onto the adjacent land. The second type is the *coastal flood*, also known as storm surges (*Green et al.*, 2025). The third type is the *pluvial flood*, often referred to as *flash flood* (*Cache et al.*, 2024; *Hofmann and Schüttrumpf*, 2021; *Busker et al.*, 2025) that can occur with extreme rainfalls. Machine Learning (ML) has become an essential element for the development of hydrological simulation and flood models (*Mosavi et al.*, 2018; *Bentivoglio et al.*, 2022). Each type of flood has unique drivers and impacts. Consequently, ML methods require different strategies to forecast them. Related tasks to flood forecasting are urban flood modeling (*Guo et al.*, 2021; *Sun et al.*, 2023; *Seleem et al.*, 2022; *Burrichter et al.*, 2023), flood inundation (*Chu et al.*, 2020; *Karim et al.*, 2023; *Pianforini et al.*, 2025), and flood extension and susceptibility mapping (*Lütjens et al.*, 2024; *Janizadeh et al.*, 2021; *Dasgupta et al.*, 2021; *Bountos et al.*, 2024). In Chapter 7, we propose a model to forecast riverine floods (fluvial) based on river discharge.

3.2.3.2 River Discharge Forecasting

River discharge can be used to detect fluvial flood signals when the magnitude of the flow exceeds certain thresholds. Current deep learning methods for forecasting river discharge are primarily based on locally lumped models (*Palash et al.*, 2024; *Zhong et al.*, 2024), hypothesizing that a single model can generalize across many catchments without considering the spatial-temporal information over grids (*Kratzert et al.*, 2024). The dominating backbone is the LSTM model (*Hochreiter and Schmidhuber*, 1997) which is used in most recent studies such as EA-LSTM *Kratzert et al.* (2019); *Heudorfer et al.* (2025), ED-LSTM (*Zhang et al.*, 2022, 2024a), Hydra-LSTM *Ruparell et al.* (2025), MC-LSTM (*Wang et al.*, 2025a), MF-LSTM (*Acuña Espinoza et al.*, 2025), and DRUM (*Ou et al.*, 2025). These models learn features specific to individual rivers or entities and lack spatial and topological information. However, river networks have spatio-temporal causal relations (*Stein et al.*, 2025). Only a few studies deviate from this conventional modeling and propose to model the network topology with Graph Neural Networks (*Kirschstein and Sun*, 2024; *Roudbari et al.*, 2024; *Kazadi et al.*, 2024). They are still limited to small scales and the graph models fail in most cases to capture topological information (*Kirschstein and Sun*, 2024). Others applied an LSTM model on a coarse grid to estimate runoff and then coupled it with a river routing model to produce daily discharge at coarse resolution (*Yang et al.*, 2025). In (*Vischer et al.*, 2025), LSTM resolves local runoff spatially on a

regular grid in central Europe. Then, routing the runoff along the entire river networks is implemented as 1D-convolutions and fully connected layers. The impact of defining routing explicitly with physics-informed neural networks has also showed an advantage in recent studies, especially, in improving streamflow in large continental river networks compared to models that do not consider routing (*Bindas et al.*, 2024; *Song et al.*, 2025). In a hybrid modeling framework, physical equations including river routing are parametrized using 3D-convolutions and fully connected layers for distributed hydrological modeling (*Wang et al.*, 2024). The most relevant work is the Encoder-Decoder LSTM (*Kratzert et al.*, 2019) developed for the Google global operational forecasting system (*Nearing et al.*, 2024), which is a locally lumped model.

In Chapter 7, distinct from previous works, we propose an approach that is capable of modeling a large spatio-temporal context and forecasting medium-range river discharge at grid-scale.

3.3 Techniques in Neural Networks

In this section, we briefly discuss previous works related to neural networks techniques that are used in this dissertation. These techniques include conditional normalization, multi-branch design, temporal positional encoding, and state space models.

3.3.1 Conditional Normalization

Since the introduction of normalization techniques in deep learning, they became a basic building block in many state-of-the-art models. Common normalization methods include batch normalization (*Ioffe and Szegedy*, 2015), group normalization (*Wu and He*, 2018), instance normalization (*Ulyanov et al.*, 2016), and layer normalization (*Ba et al.*, 2016). It has been shown empirically that normalization layers help with model optimization and regularization. Through normalization layers, the activation maps inside the model are normalized to follow a normal distribution with zero mean. After that, the normalized activation maps are modulated or denormalized by learnable affine transformation parameters. These parameters vary across channels and are learned based on the running training statistics together with the model parameters. Therefore, such a normalization method is called unconditional. Compared to popular unconditional normalizations, there exist conditional normalization techniques which aim to learn affine parameters conditionally on external input. In the field of computer vision, conditional normalization is often used for image synthesis and style transformation (*Park et al.*, 2019; *Wang et al.*, 2018; *Ling et al.*, 2021; *van Noord and Postma*, 2017; *Zhu et al.*, 2020; *Tan et al.*, 2022; *Lv et al.*, 2021b; *Lyu et al.*, 2021; *Jakoel et al.*, 2022; *Chen et al.*, 2019; *Huang and Belongie*, 2017; *Sushko et al.*, 2022; *Li et al.*, 2022b). More recently, *Marín and Escalera* (2021) adapted the conditional normalization from *Wang et al.* (2018) and *Park et al.* (2019) to generate high-resolution satellite images.

In Chapters 4 and 7, we use the conditional normalization in a very different way than *Marín and Escalera* (2021), *Wang et al.* (2018), and *Park et al.* (2019). While these works focus on synthesizing an image using the segmentation map as conditional input, we aim to modulate dynamic features conditioned on static features.

3.3.2 Multi-Branch Neural Networks

When deep learning is applied to potentially multi-source remote sensing-based Earth observation data, multi-branch neural networks are a commonly used framework. This is mainly because such networks enrich representation learning and provide discriminative learning perspectives of the input variables (*Zhao et al.*, 2022). In addition, an important aspect of the multi-branch design is the capability to adapt some parts of the model to a specific type of input. The general framework generates features from each branch and fuses these features in the network to obtain a unified feature vector. This fused representation is used as input to the subsequent layers. In *Gaetano et al.* (2018), a two-branch 2D CNN network was proposed to handle panchromatic information along with a multi-spectral one for image classification. *Tan et al.* (2020) reduced the depth of a semantic segmentation classifier by applying consecutive blocks, each containing three CNN branches. A similar objective can be found in *Zhao et al.* (2022), where the network complexity was reduced via weight sharing and self-distillation embedding. In this way, only the main branch is used during inference, which inherits the knowledge of trained subbranches and has a close performance to an ensemble model. For hyperspectral image classification, *Xu et al.* (2018) introduced a model called Spectral-Spatial Unified Network (SSUN). In their model, spectral features are learned by a grouping-based LSTM, and spatial features are learned by a 2D CNN. *Shen et al.* (2019) used separated spectral and spatial convolutional branches for hyperspectral input (S^2CDELM). A multi-branch architecture was also explored for image fusion. *Liu et al.* (2019) proposed a two-stream CNN called (StfNet). They investigated the task of spatio-temporal image fusion. Their network takes a coarse image input along with its neighboring images to predict the reconstructed fine image. Some works adapted a multi-branch architecture to construct a multi-scale feature vector. In *Gan et al.* (2020), a dual-branch CNN with different filter kernel sizes was used as an autoencoder. Thus, the input image could be processed at different scales. *Tang et al.* (2022b) proposed a multi-scale Gaussian pyramid to handle hyperspectral input. They used the Gaussian pyramid to obtain multi-scale images which are then processed by ResNet modules (*Zhong et al.*, 2017). In this way, spatial features can be learned at different scales. They further used a second branch, which performs a discrete wavelet transform on the spectral input followed by an LSTM module (*Hochreiter and Schmidhuber*, 1997). The spatial and spectral features are then fused and processed by an MLP to obtain the final classification result. For short-term multi-temporal image classification, *Zheng et al.* (2021) addressed the task through a Multi-temporal Deep Fusion Network (MDFN). In their framework, the LSTM-based branch is used to learn temporal-spectral features. At the same time, temporal-spatial and spectral-spatial information is learned via a joint 3D-2D CNN with two branches. Furthermore, some works employed attention mechanisms with multi-branch architectures (*Li et al.*, 2020; *Zhu et al.*, 2022; *Deng et al.*, 2022). When attention is applied, it drives the model to focus more on regions of interest. The former described methods (*Gaetano et al.*, 2018; *Tan et al.*, 2020; *Xu et al.*, 2018; *Shen et al.*, 2019; *Liu et al.*, 2019; *Gan et al.*, 2020; *Tang et al.*, 2022b; *Zheng et al.*, 2021; *Li et al.*, 2020) except of *Zhao et al.* (2022), *Zhu et al.* (2022), and *Deng et al.* (2022) did not consider linking information between branches during the feature learning stage. This limits the gradient flow and disentangles the correlations between the learned features.

In Chapter 4, we propose an architecture composed of two CNN branches for a forecasting task. A 2D branch is used to learn spatial features from static variables. At the same time, a 3D branch is used to learn spatio-temporal features from dynamic variables, which vary along the input temporal

dimension. The branches are further linked via adaptive modulation layers to model the causal effects of static variables on dynamic variables. In Chapter 6, we propose a model to identify anomalies in the weather data utilizing the multi-branch architecture.

3.3.3 Temporal Positional Encoding

A plethora of studies exist about temporal modeling in remote sensing (*Shi et al.*, 2015; *Rußwurm and Körner*, 2018; *Pelletier et al.*, 2019; *Moskolaï et al.*, 2021). Recently, Transformer with self-attention, as it is presented by *Vaswani et al.* (2017) for NLP has become a natural choice to handle sequential data, which includes a positional encoding. In the field of remote sensing, many works showed the benefits of adapting positional encoding for time-dependent image classification (*Garnot et al.*, 2020; *Garnot and Landrieu*, 2020), panoptic segmentation (*Fare Garnot and Landrieu*, 2021), and image generation (*Drees et al.*, 2022). In their work, each image was given an encoded time vector according to its position with respect to a reference point, e.g., the first acquisition time step. *Nyborg et al.* (2022b) used the calendar time (day of the year) to provide positional information within the sequence. They also proposed to learn or estimate time shifts between geographically distant regions to enhance the generalization further. In another work, *Nyborg et al.* (2022a) used the thermal time, which is obtained by accumulating daily average temperatures over the growing season, for crop classification. In general, transformer-based approaches require the positional encoding since the temporal information is otherwise lost within a transformer model (Section 2.1.5).

While 3D CNNs consider the temporal order of the input such that a positional encoding is not necessary, we show in Chapter 4 that adding an absolute temporal encoding is also useful for time-dependent forecasting.

3.3.4 State Space Model (SSM) and Mamba

Building upon Mamba (Section 2.1.7), VMamba (*Liu et al.*, 2024) and Vim (*Zhu et al.*, 2024) in the vision domain were introduced as appealing alternatives to the quadratic complexity of ViTs (Section 2.1.5) while improving scaling efficiency on long token sequences. A series of works have adapted Mamba to tasks like image generation (*Phung et al.*, 2024; *Hu et al.*, 2025), image classification (*Pei et al.*, 2024; *Huang et al.*, 2025; *Xiao et al.*, 2024; *Shaker et al.*, 2025), video understanding (*Li et al.*, 2025; *Chen et al.*, 2024), motion generation (*Zhang et al.*, 2025), dense action anticipation (*Zatsarynna et al.*, 2025), and point cloud processing (*Liang et al.*, 2024; *Zhang et al.*, 2024b). Mamba and SSMs were also applied to local-scale runoff prediction as a time-series forecast (*Jia et al.*, 2024; *Wang et al.*, 2025b).

In Chapter 7, we use SSM for flood forecasting, where we model river points as one sequence and apply Mamba to connect the river networks at a large scale.

CHAPTER 4

Location-aware Adaptive Normalization: A Deep Learning Approach For Wildfire Danger Forecasting

---

This chapter is based on the following publication (*Shams Eddin et al.*, 2023):

Location-aware Adaptive Normalization: A Deep Learning Approach For Wildfire Danger Forecasting
Mohamad Hakam Shams Eddin, Ribana Roscher, and Juergen Gall
IEEE Transactions on Geoscience and Remote Sensing (TGRS), 2023
Doi: 10.1109/TGRS.2023.3285401

Below we summarize the contributions of each author.

- **Mohamad Hakam Shams Eddin**
  Proposed the task of wildfire forecasting and the main idea of the paper. Hakam designed the method, the experimental setup, conducted the experiments, preprocessed the data, implemented the codebase, and wrote the initial manuscript.

- **Ribana Roscher**
  Contributed with discussions and writing.

- **Juergen Gall**
  Proposed the idea of using adaptive normalization. Juergen supervised and funded the project. In addition, he contributed with discussion and writing.

Contents

4.1 Introduction . . . 52
4.2 Methodology . . . 53
4.2.1 2D/3D Two-Branch CNN . . . 54
4.2.2 Location-Aware Adaptive Normalization (LOAN) . . . 55
4.2.3 Absolute Temporal Encoding (TE) . . . 58
4.3 Experimental Results and Analysis . . . 58
4.3.1 Experimental Setup . . . 58
4.3.2 Comparison with Baselines . . . 60
4.3.3 Qualitative Results . . . 63

4.3.4 Variable Importance 71
4.3.5 Comparing Different Conditional Maps 71
4.3.6 Ablation Studies 72
4.4 Summary 75

In this chapter, we address the problem of forecasting wildfire danger for the next day. The forecasting model starts with spatio-temporal data as an initial condition and predicts for the next day the susceptibility map of wildfire danger. In the context of wildfire danger forecasting, previous machine learning approaches duplicate static variables along the time dimension and input static and dynamic variables simultaneously to the model. In this chapter, we argue that this neglects the inherent differences between static and dynamic input variables. Furthermore, most existing multi-branch architectures lose the interconnections between the branches during the feature learning stage. To address these issues, in this chapter, we proposes a 2D/3D two-branch convolutional neural network (CNN) with a Location-aware Adaptive Normalization layer (LOAN). Using LOAN as a building block, we can modulate the dynamic features conditional on their geographical locations. Thus, our approach considers feature properties as a unified yet compound 2D/3D model. Besides, we propose using the sinusoidal-based encoding of the day of the year to provide the model with explicit temporal information about the target day within the year. Our experimental results show a better performance of our approach than other baselines on the challenging FireCube dataset. The results show that location-aware adaptive feature normalization is a promising technique to learn the relation between dynamic variables and their geographic locations, which is highly relevant for areas where remote sensing data builds the basis for analysis.

4.1 Introduction

There is a general expectation that weather and climate extremes will change their patterns and frequencies in the future (*Ren et al.*, 2013; *Lausier and Jain*, 2018; *Perkins-Kirkpatrick and Lewis*, 2020; *Samuels et al.*, 2018). This is particularly the case for the Mediterranean region, which has been identified as a hot spot for climatic changes (*Zittis et al.*, 2019; *Barcikowska et al.*, 2020; *Hochman et al.*, 2022). Because extreme weather events can impose short- and long-term risks in our Earth system, predicting these risks such as droughts, windstorms, and wildfires has become recently more relevant. In particular, wildfire forecasting constitutes one of the open challenges for risk assessment and emergency response (*Thompson et al.*, 2019; *Coogan et al.*, 2019; *Moreira et al.*, 2020). Wildfire forecasting refers to the task of fire-susceptibility mapping using key remote sensing, meteorological, and anthropogenic variables (*Jain et al.*, 2020). Building an integrated modeling system of the Earth should also consider wildfire events to comprehend the origin of past patterns better and predict future ones (*Fornacca et al.*, 2017). Unlike typical prediction tasks, understanding when weather conditions have a high tendency to cause fire events addresses more complexities; among these are the stochastic nature of fire events (*Kondylatos et al.*, 2022) and fire drivers, which are time-dependent and intercorrelated across variables (*Hantson et al.*, 2016). Moreover, the prediction model should consider difficulties like a high false positive error rate, uncertainty, and class imbalance.

In recent years, many works leveraged classical machine learning approaches to solve the task (*Jain et al.*, 2020). More recently, deep learning methods (*Reichstein et al.*, 2019) have become

popular since they can handle large multivariate datasets more efficiently and are able to learn highly complex relationships between observations and the predicted outcome. In the context of wildfire danger forecasting, *Prapas et al.* (2021) and *Kondylatos et al.* (2022) proposed to use recurrent neural networks in combination with 2D convolutions to exploit both temporal and spatial context. These approaches, however, do not distinguish between the different input variables. Static variables like elevation, which barely change over time, are simply copied and concatenated with dynamic variables like surface temperature. This results not only in a highly redundant input to the network, but it also neglects strong causal effects between static and dynamic variables. For instance, the surface temperature strongly depends on the geographical location, which is described by static variables.

In this work, we thus propose a convolutional neural network for wildfire danger forecasting that handles static and dynamic variables differently. Since the static variables do not change over time, they are processed by a branch consisting of 2D convolutions while the dynamic variables are processed by the second branch with 3D convolutions as illustrated in Fig. 4.1. To address the causal effect of static variables on dynamic variables, we introduce *feature modulation* for the dynamic variables where the modulation parameters are generated dynamically and conditionally on the geographical location. We thus name this method Location-aware Adaptive Normalization (LOAN). In addition, we encode the date of the forecasting during a year by an absolute time encoding based on the sinusoidal encoding (*Vaswani et al.*, 2017). Both LOAN and the time encoding can be implemented as plugin layers in different deep learning architectures. We view our model as a generic architecture that can be used for other time-dependent forecasting tasks with static and dynamic variables. We conduct extensive experiments on the FireCube dataset (Section 2.2.1) where our approach outperforms previous works. We achieve an overall improvement of up to 5.72% in precision, 3.24% in F1-score, 0.63% in AUROC, and 1.15% in OA on the test set.

The rest of this chapter is organized as follows. Section 4.2 describes the proposed method is described in detail. The experimental results and ablation study are provided in Section 4.3. Finally, a summary is given in Section 4.4.

4.2 Methodology

Problem formulation. Given a multivariate spatio-temporal data cube $\mathcal{C}(\{V, T, W, H\})$, where H and W are the spatial extensions of the cube, T is the temporal extension in the past for the time series $1, 2, \ldots, T$, and V is the number of variables (static and dynamic), our aim is to learn a mapping function f approximated by a neural network that can predict the probability $Y_{T+1} \in [0, 1]$ of a wildfire event to start at the center of $W \times H$ for the target day $T+1$:

$$f : \mathcal{C}(\{V, T, W, H\}) \rightarrow Y_{T+1}\,. \tag{4.1}$$

To achieve this, we propose a spatio-temporal 2D/3D CNN with two branches as illustrated in Fig. 4.1. First, the network design is introduced in Section 4.2.1. Then, the Location-aware Adaptive Normalization layer (LOAN), which is the core of our work is explained in detail in Section 4.2.2. Finally, Section 4.2.3 describes the integration of the absolute temporal encoding (TE) into the model.

4.2.1 2D/3D Two-Branch CNN

As shown in Fig. 4.1, our network consists of two branches that process dynamic and static variables, respectively. We denote the data cube with dynamic variables by $\mathcal{C}(\{V_d, T, W, H\})$ and the data cube with static variables by $\mathcal{C}(\{V_s, W, H\})$. As in previous works, we normalize the input channel-wise to the range $[0, 1]$. Since the static variables do not have a time component, we use 2D convolutions for the static branch and 3D convolutions for the dynamic branch. More in detail:

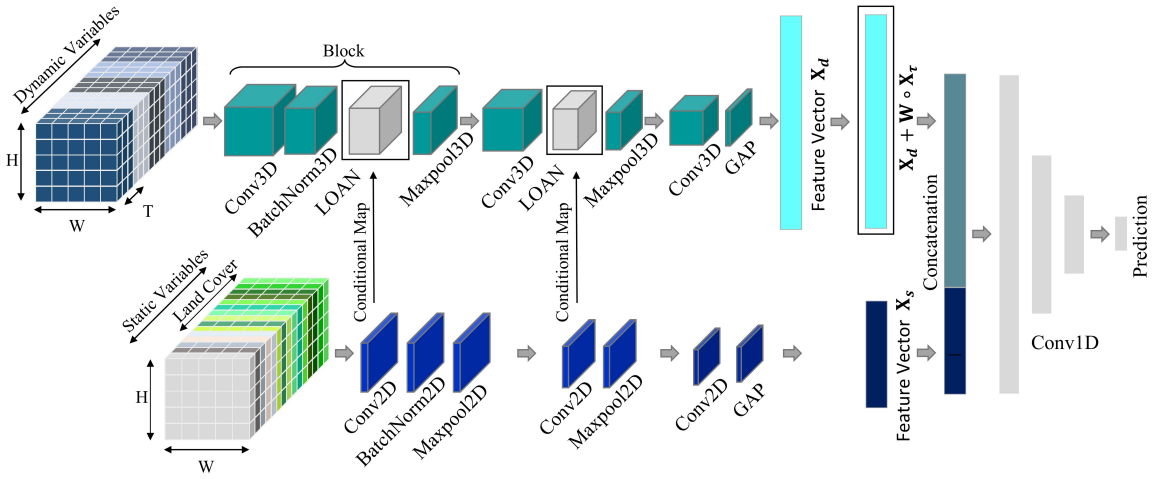


Figure 4.1: Overview of the proposed approach. Our network handles static and dynamic variables by two branches. For the static branch, we use 2D convolutions whereas we use 3D convolutions for the dynamic branch since the dynamic variables change over time. Since dynamic variables like surface temperature strongly depend on static variables like elevation, we modulate the dynamic features conditioned on the static features at two blocks (LOAN). This makes the model location-aware since the static variables contain geographic data like land cover. The feature vector $\mathbf{X}_d$ from the dynamic branch is combined with a weighted temporal encoding vector $\mathbf{X}_\tau$ before it is concatenated with the static feature vector $\mathbf{X}_s$. The concatenated vector serves as input to fully connected layers that predict the probability of a wildfire.

Dynamic branch. The dynamic branch takes the variables V_d which vary over time as input. It consists of 3 blocks; each block has a 3D convolution with a 3×3 kernel size followed by a ReLU activation function and a 3D max pooling layer. To reduce overfitting, we use global average pooling (GAP) (*Lin et al.*, 2013) at the end of the last block. We denote the feature vector learned from this branch as $\mathbf{X}_d \in \mathbb{R}^{256}$.

Static branch. In parallel to the dynamic branch, the static branch has a similar architecture. However, 2D convolutions are used instead of 3D ones. We denote the feature vector learned from this branch as $\mathbf{X}_s \in \mathbb{R}^{128}$. Note that the dimensionality of the static feature vector is lower than the dimensionality of the dynamic feature vector since the input data cube is smaller.

In a nutshell, the dynamic- and static branch are two functions f_d and f_s, respectively:

$$f_d : \mathcal{C}(\{V_d, T, W, H\}) \rightarrow \mathbf{X}_d\,, \tag{4.2}$$

$$f_s : \mathcal{C}(\{V_s, W, H\}) \rightarrow \mathbf{X}_s\,. \tag{4.3}$$

For the dynamic feature vector we add an absolute temporal encoding $\mathbf{X}_\tau$, which will be described in Section 4.2.3. The two feature vectors are then concatenated and fed into 4 classification layers with 1D convolutions of kernel size 1. The layers reduce the dimensionality from 384 to 256, 128, 32, and 2. To reduce overfitting, we use dropout with a dropout probability $p = 0.5$ for the 1D convolutional layers except the last two layers. Finally, a softmax activation is used after the last classification layer to predict the probability of a wildfire. In addition, we use a batch normalization layer (*Ioffe and Szegedy*, 2015) for the 1^{st} block of each branch. More implementation details are given in Section 4.3.

For training, we use the binary cross entropy as loss function:

$$-\frac{1}{N}\sum_{n=1}^{N}\left[\hat{Y}_{T+1}^{(n)}\log(Y_{T+1}^{(n)})+(1-\hat{Y}_{T+1}^{(n)})\log(1-Y_{T+1}^{(n)})\right], \tag{4.4}$$

where N denotes the batch size and $\hat{Y}_{T+1}^{(n)} \in \{0,1\}$ is the true label for sample n.

In the following, we discuss the Location-Aware Adaptive Normalization (LOAN) that modulates the dynamic features based on the static features and the already mentioned absolute temporal encoding.

4.2.2 Location-Aware Adaptive Normalization (LOAN)

In general, dynamic variables are correlated with the geographic location, e.g., temperature and pressure change with elevation, soil moisture and NDVI vary with land cover, and humidity is correlated with some static variables like the waterway distance. Since the dynamic variables depend on the static variables and not vice versa, we aim to exploit this knowledge in our approach. This is done by learning to normalize the dynamic features based on the location-dependent static features. To this end, we propose a conditional normalization technique for remote sensing data called *Location-aware Adaptive Normalization (LOAN)*.

We first describe a batch-normalization (*Ioffe and Szegedy*, 2015) where the activation map is normalized before it is modulated by scale γ and bias β. Let ${z_d}^i \in \mathbb{R}^{N\times K^i\times D^i\times W^i\times H^i}$ be an activation map in the i-th block of the dynamic branch and ${z_s}^i \in \mathbb{R}^{N\times K^i\times W^i\times H^i}$ be an activation map of the corresponding i-th block in the static branch, where K^i denotes the number of channels and D^i, W^i and H^i denote the depth, width, and height of the activation map z^i, respectively. Using the indices $n \in \{1,\dots,N\}$, $k \in \{1,\dots,K^i\}$, $t \in \{1,\dots,D^i\}$, $w \in \{1,\dots,W^i\}$, and $h \in \{1,\dots,H^i\}$, the normalization of the dynamic branch is performed by the following equation:

$$\left(\frac{{z_d}^i_{(n,k,t,w,h)}-\mu_k}{\sigma_k}\right)\cdot\gamma_k+\beta_k\,, \tag{4.5}$$

where ${z_d}^i_{(n,k,t,w,h)}$ is the activation before the normalization, μ_k and σ_k are the computed mean and standard deviation of channel k, e.g., computed over the tensor $D^i \times W^i \times H^i$ and all samples n in the batch, and γ_k and β_k are the learnable modulation parameters.

In our case, we aim to learn a modulation of the dynamic features ${z_d}^i_{(n,k,t,w,h)}$ at the i-th block where the modulation parameters ${\gamma_s}^i_{(n,k,w,h)}$ and ${\beta_s}^i_{(n,k,w,h)}$ depend on the corresponding static features ${z_s}^i_{(n,k,w,h)}$:

$${z_d}^i_{(n,k,t,w,h)}\cdot{\gamma_s}^i_{(n,k,w,h)}+{\beta_s}^i_{(n,k,w,h)}\,. \tag{4.6}$$

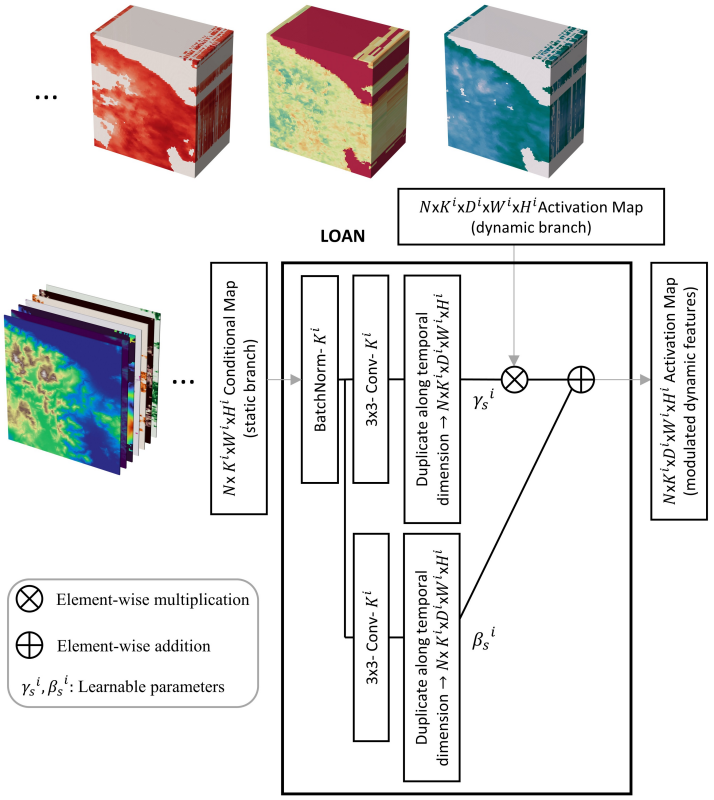


Figure 4.2: Illustration of the Location-aware Adaptive Normalization layer (LOAN) when using the conditional map from the static branch. The conditional map of the static branch and the activation map of the dynamic branch have the same N, K^i, W^i, and H^i dimensions. BatchNorm denotes a batch normalization layer. 3×3-Conv-K^i denotes a convolution layer with a kernel size of 3×3 and K^i output channels.

In contrast to (4.5), the modulation parameters ${\gamma_s}^i_{(n,k,w,h)}$ and ${\beta_s}^i_{(n,k,w,h)}$ vary with respect to sample n in the batch, the location (w, h), and across channel k, but they are constant over time t. Furthermore, they are conditioned on the static features ${z_s}^i$. We thus call ${z_s}^i$ the conditional map for the modulation.

The way how ${\gamma_s}^i_{(n,k,w,h)}$ and ${\beta_s}^i_{(n,k,w,h)}$ are computed is illustrated in Fig. 4.2. First, the conditional map ${z_s}^i$ is normalized channel-wise as following:

$$\hat{z_s}^i_{(n,k,w,h)} = \frac{{z_s}^i_{(n,k,w,h)} - {\mu_k}^i}{{\sigma_k}^i}, \tag{4.7}$$

where

$${\mu_k}^i = \frac{1}{NW^iH^i} \sum_{n=1}^{N} \sum_{w=1}^{W^i} \sum_{h=1}^{H^i} {z_s}^i_{(n,k,w,h)}, \tag{4.8}$$

$${\sigma_k}^i = \sqrt{\frac{1}{NW^iH^i} \sum_{n=1}^{N} \sum_{w=1}^{W^i} \sum_{h=1}^{H^i} ({z_s}^i_{(n,k,w,h)} - {\mu_k}^i)^2}. \tag{4.9}$$

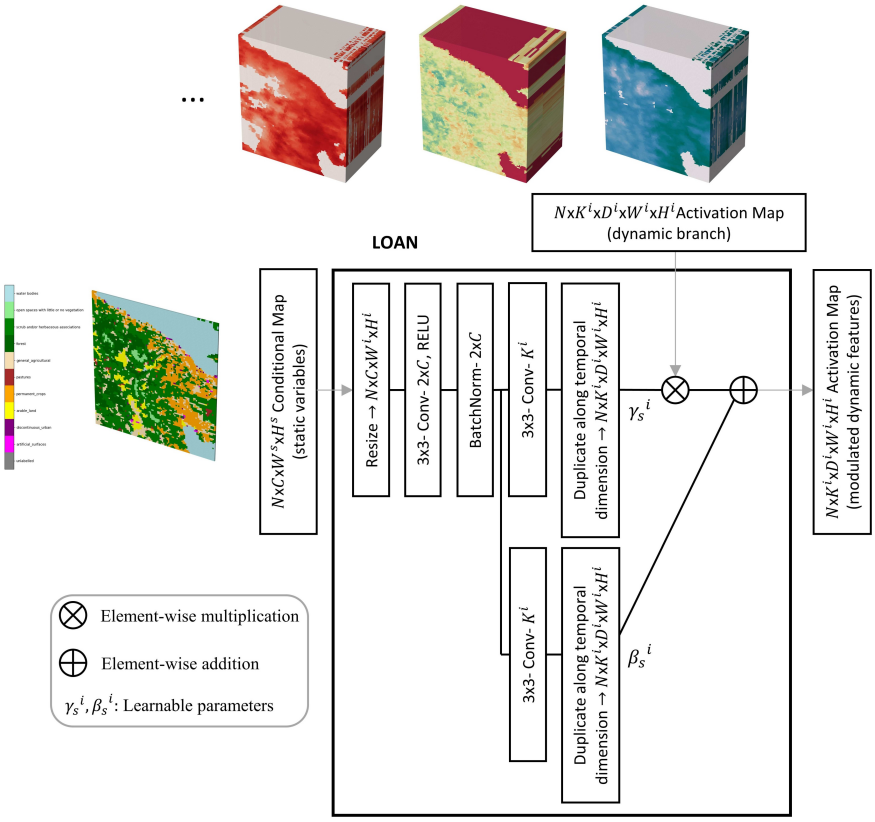


Figure 4.3: Illustration of the Location-aware Adaptive Normalization layer (LOAN) when using C static variables directly for modulation. In this case, there is a mismatch between the dimensions of the conditional map and the activation maps of the dynamic branch at the i-th block. To adjust for the spatial resolution, we resize the conditional map using nearest-neighbor down-sampling to match the resolution of the activation map from the dynamic branch. The convolution layer takes C channels as input and generates an output with K^i channels.

Afterward, $\hat{z_s}^i_{(n,k,w,h)}$ is projected by two convolutional layers with K^i filters to compute $\gamma_s{}^i_{(n,k,w,h)}$ and $\beta_s{}^i_{(n,k,w,h)}$. In our implementation, these modulation parameters are then duplicated along the temporal dimension to match the depth D^i of $z_d{}^i$ such that (4.6) can be computed.

We add the Location-aware Adaptive Normalization layer (LOAN) in the first two blocks as shown in Fig. 4.1. The activation maps of the dynamic branch are normalized only in the 1st block and modulated in both the 1st and 2nd blocks. The impact of the blocks where the LOAN layer is added is evaluated in Table 4.7.

In the experimental section, we also evaluate a variant of LOAN that is not conditioned on the intermediate features of the static branch as shown in Figs. 4.1 and 4.2, but on the static variables directly as shown in Fig. 4.3. In this case, the conditional map has different spatial dimensions and number of channels compared to the features in the dynamic branch, e.g., the conditional maps consist of C variables and have $W^s \times H^s$ spatial dimensions. In this respect, the conditional map ($W^s \times H^s$) is first resized to match the spatial dimensions ($W^i \times H^i$) of the activation map from the dynamic branch. We use the nearest-neighbor method for the down-sampling. The resized con-

ditional map is then fed into a convolutional layer with 3×3 kernel size to double the number of channels, e.g., $2 \times C$. Finally, as in the previous version of LOAN, the conditional map is normalized, projected by two convolutional layers, and duplicated along the temporal dimension to compute $\gamma_{s(n,k,w,h)}^{i}$ and $\beta_{s(n,k,w,h)}^{i}$. The impact of different conditional maps using the variants of LOAN is evaluated in Table 4.6.

4.2.3 Absolute Temporal Encoding (TE)

Some extreme events in the climate model have a dependent relation on time (*Martín et al.*, 2019; *Hantson et al.*, 2016). This is also the case for the FireCube dataset (Section 2.2.1) where wildfire events vary from month to month and occur more frequently in the summer time as shown in Fig. 4.5. So far, the network does not consider an absolute time like the day of the year. Instead, for any forecast day $T+1$, the last 10 days are used as observations but the network does not have the information what day during the year T is.

As shown in Fig. 4.1, we add this information to the dynamic branch before we concatenate the static and dynamic features. To encode the day of the year, we use the standard fixed sinusoidal-based encoding by *Vaswani et al.* (2017). We pre-compute for each day of the year $\tau \in [0, 365]$[1], which is extracted from T, the absolute temporal encoding vector $\mathbf{X}_\tau \in \mathbb{R}^{256}$:

$$\mathbf{X}_\tau(2j) = \sin(\tau/10^{2j/256}), \tag{4.10}$$

$$\mathbf{X}_\tau(2j+1) = \cos(\tau/10^{2j/256}), \tag{4.11}$$

where j is the embedding dimension. Each even dimension results from a sine function, while each odd dimension results from a cosine function. This allows τ to have a smooth and yet unique encoding for every time step, e.g., each day of a year. Note that the vector has the same size as $\mathbf{X}_d$.

In order to add the absolute time encoding vector $\mathbf{X}_\tau$ to the dynamic feature vector $\mathbf{X}_d$, we weight each element of the vector by a learnable weight vector $\mathbf{W}_\tau \in \mathbb{R}^{256}$:

$$\mathbf{X}_d + \mathbf{W}_\tau \odot \mathbf{X}_\tau, \tag{4.12}$$

where $\odot$ denotes the Hadamard product, e.g., element-wise multiplication. Fig. 4.4 illustrates how the temporal embedding is added to the model.

4.3 Experimental Results and Analysis

4.3.1 Experimental Setup

4.3.1.1 Implementation Details

The network is trained with the binary cross entropy loss (4.4) using the Pytorch framework (*Paszke et al.*, 2019) with a learning rate 0.00003 and the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$) (*Kingma and Ba*, 2014) with a weight decay 0.02. We use a batch size of $N = 256$ and train the network for 40 epochs. All models were trained on a single NVIDIA GeForce RTX 3090 GPU. Additional details and code are available at `https://github.com/HakamShams/LOAN`.

[1] We consider February 29 for the encoding.

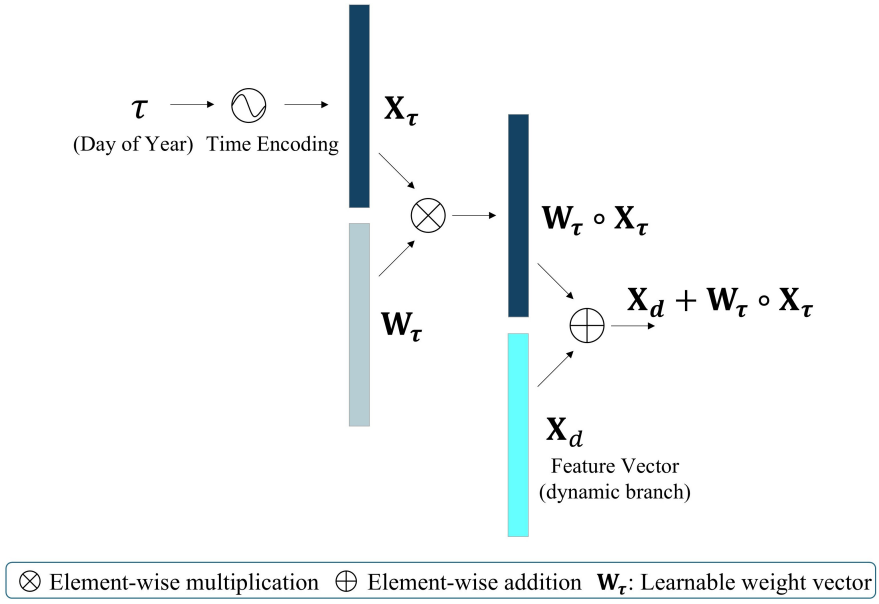


Figure 4.4: Illustration of how the absolute temporal encoding is added in the model. The day of year τ is encoded into a vector $\mathbf{X}_\tau$ and each element is weighted by the learned weight vector $\mathbf{W}_\tau$. The weighted vector is then added to the dynamic feature vector $\mathbf{X}_d$.

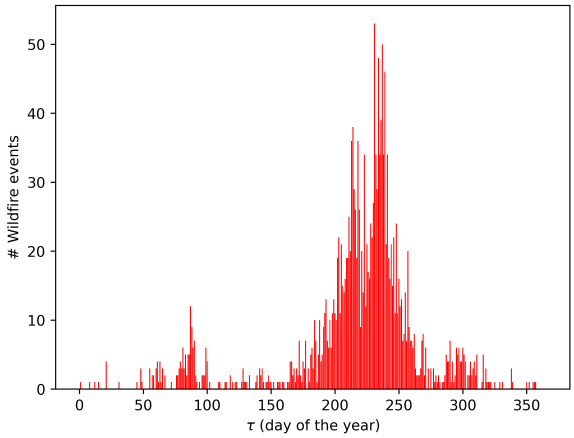


Figure 4.5: Distribution of the wildfire events per day between 06/03/2009 and 29/08/2021 (FireCube dataset).

4.3.1.2 Baselines

We compare our approach to the approaches that have been evaluated on the described dataset in *Kondylatos et al.* (2022). This includes two deep learning models, namely LSTM (*Hochreiter and Schmidhuber*, 1997) and ConvLSTM (*Shi et al.*, 2015), and two classical machine learning classifiers, namely Random Forests (RF) (*Breiman*, 2001) and XGBoost (*Chen and Guestrin*, 2016). For further details regarding the architectures and hyper-parameters of the models, we refer to the work of *Kondylatos et al.* (2022). In order to demonstrate the benefit of treating static and dynamic variables differently, we also compare with a one-branch 3D CNN where we duplicate the static variables

along the temporal dimension and concatenate them together with the dynamic variables to form a single data cube.

In addition, we compare to the recent transformer models TimeSformer (*Bertasius et al.*, 2021) and Video SwinTransformer (*Liu et al.*, 2022b), which use space-time attention. Vision transformers are based on a self-attention mechanism to model spatio-temporal dependencies. Compared to CNNs, transformers have less inductive bias and need much more computational resources for training. To the best of our knowledge, no prior work has done a systematic study about the performance of video vision transformers for wildfire forecasting. To ensure a fair comparison, all baselines were re-implemented and trained on the same samples with a fixed random seed. We do not use any augmentation technique.

As a tradional baseline for wildfire danger forecasting, we additionally compare to the Fire Weather Index (FWI) (*Van Wagner et al.*, 1985) in Section 4.3.3.

4.3.1.3 Datasets

As described in Section 4.1, we use the FireCube dataset (Section 2.2.1).

4.3.1.4 Evaluation Metrics

We follow the same protocol for quantitative comparison as in *Prapas et al.* (2021) and *Kondylatos et al.* (2022). The evaluation metrics are precision, recall, and F1-score, calculated for the positive class that represents a wildfire event. In addition, we report true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). Moreover, we provide the overall accuracy (OA) and the area under the receiver operating characteristic curve (AUROC) as evaluation metrics.

4.3.2 Comparison with Baselines

The quantitative results of our experiments are provided in Table 4.1. The results of the proposed 2D/3D CNN are shown with and without absolute temporal encoding (TE).

We can observe that the proposed 2D/3D CNN outperforms the other methods for most metrics, particularly FP, TN, Precision, F1-score, and OA, on the validation and testing sets. SwinTransformer and ConvLSTM achieve a slightly higher AUROC for 2019 and 2020, respectively. LSTM and SwinTransformer achieve a higher recall, but at the cost of a very low precision. In comparison with other deep learning methods, LSTM and SwinTransformer have even the highest number of false positives for all years. The main weakness of LSTM lies in the fact that it does not consider the spatial context, while large models like SwinTransformer are prone to overfitting, which results in a relatively poor performance for 2020. RF and XGBoost have the same disadvantage as LSTM, but even a weaker temporal model and thus perform worse than LSTM. Most interesting is the comparison to 3D CNN since it uses the same 3D CNN structure but only one branch, e.g., it treats static variables like dynamic variables. The results show that the proposed approach with two branches outperforms the single-branch architecture for all metrics and all years. This demonstrates the importance of treating static variables differently than dynamic variables. Adding the absolute temporal encoding (TE) to the model substantially reduces FP at the cost of decreasing TP. This is also reflected in the precision and recall.

In Tables 4.2 and 4.3, we present additional experimental results alongside the memory footprint as the number of parameters (# Params), the estimated multiply-accumulate operations (multiply-adds) (MMACs), and the expected inference time, which is estimated as samples per millisecond (# SPmS). The performance metrics are calculated on both testing years 2020-2021 as one set. Since

Year 2019 (val) - 6430 samples									
Algorithm	TP(↑)	FP(↓)	TN(↑)	FN(↓)	Precision(↑)	Recall(↑)	F1(↑)	AUROC(↑)	OA(↑)
RF (*Kondylatos et al.*, 2022)	575	372	4758	725	0.607	0.442	0.512	0.885	0.829
XGBoost (*Kondylatos et al.*, 2022)	928	448	4682	372	0.674	0.714	0.694	0.923	0.873
LSTM (*Kondylatos et al.*, 2022)	968	431	4699	332	0.692	0.745	0.717	0.936	0.881
ConvLSTM (*Kondylatos et al.*, 2022)	867	276	4854	433	0.759	0.667	0.710	0.947	0.890
TimeSformer (*Bertasius et al.*, 2021)	967	248	4882	333	0.796	0.744	0.769	**0.957**	0.910
SwinTransformer (*Liu et al.*, 2022b)	**979**	324	4806	**321**	0.751	**0.753**	0.752	0.945	0.900
3D CNN	918	265	4865	382	0.776	0.706	0.739	0.942	0.899
2D/3D CNN (ours)	970	248	4882	330	0.796	0.746	**0.771**	0.945	0.910
2D/3D CNN w/ TE (ours)	905	**182**	**4948**	395	**0.833**	0.696	0.758	0.951	**0.910**
Year 2020 (test) - 6088 samples									
Algorithm	TP(↑)	FP(↓)	TN(↑)	FN(↓)	Precision(↑)	Recall(↑)	F1(↑)	AUROC(↑)	OA(↑)
RF (*Kondylatos et al.*, 2022)	750	245	4615	478	0.754	0.611	0.675	0.912	0.881
XGBoost (*Kondylatos et al.*, 2022)	**891**	322	4538	**337**	0.735	**0.726**	0.730	0.911	0.892
LSTM (*Kondylatos et al.*, 2022)	**891**	290	4570	**337**	0.754	**0.726**	0.740	0.936	0.897
ConvLSTM (*Kondylatos et al.*, 2022)	811	155	4705	417	0.840	0.660	0.739	**0.943**	0.906
TimeSformer (*Bertasius et al.*, 2021)	751	140	4720	477	0.843	0.612	0.709	0.924	0.899
SwinTransformer (*Liu et al.*, 2022b)	794	202	4658	434	0.797	0.647	0.714	0.928	0.896
3D CNN	797	160	4700	431	0.833	0.649	0.730	0.931	0.903
2D/3D CNN (ours)	841	160	4700	387	0.840	0.685	**0.755**	0.940	**0.910**
2D/3D CNN w/ TE (ours)	776	**117**	**4743**	452	**0.869**	0.632	0.732	0.942	0.907
Year 2021 (test) - 36732 samples									
Algorithm	TP(↑)	FP(↓)	TN(↑)	FN(↓)	Precision(↑)	Recall(↑)	F1(↑)	AUROC(↑)	OA(↑)
RF (*Kondylatos et al.*, 2022)	3264	1157	31168	1143	0.738	0.741	0.740	0.968	0.937
XGBoost (*Kondylatos et al.*, 2022)	3016	1345	30980	1391	0.692	0.684	0.688	0.959	0.926
LSTM (*Kondylatos et al.*, 2022)	3739	1359	30966	668	0.733	0.848	0.787	0.971	0.945
ConvLSTM (*Kondylatos et al.*, 2022)	3514	769	31556	893	0.821	0.797	0.809	0.978	0.955
TimeSformer (*Bertasius et al.*, 2021)	3578	867	31458	829	0.805	0.812	0.808	0.977	0.954
SwinTransformer (*Liu et al.*, 2022b)	**3962**	954	31371	**445**	0.806	**0.899**	0.850	0.981	0.962
3D CNN	3766	810	31515	641	0.823	0.855	0.839	0.980	0.961
2D/3D CNN (ours)	3870	757	31568	537	0.836	0.878	0.857	0.982	0.965
2D/3D CNN w/ TE (ours)	3841	**416**	**31909**	566	**0.902**	0.872	**0.887**	**0.985**	**0.973**

Table 4.1: Comparison with baselines. The classification metrics are shown for the years 2019-2021. The values of Precision, Recall, F1-score, OA, and AUROC are given in percent (%). (TE) denotes the absolute temporal encoding. The best and second-best results of each metric are highlighted in bold and italic text, respectively.

Year 2020-2021 (test)								
Algorithm	LOAN	TE	Params(↓)	Precision(↑)	Recall(↑)	F1(↑)	AUROC(↑)	OA(↑)
LSTM (*Kondylatos et al.*, 2022)	✗	✗	30k	0.737	0.822	0.777	0.965	0.938
ConvLSTM (*Kondylatos et al.*, 2022)	✗	✗	372k	0.824	0.768	0.795	0.971	0.948
TimeSformer (*Bertasius et al.*, 2021)	✗	✗	1.16m	0.811	0.768	0.789	0.966	0.946
SwinTransformer (*Liu et al.*, 2022b)	✗	✗	1.78m	0.805	**0.844**	0.824	0.971	0.953
3D CNN	✗	✗	323k	0.839	0.775	0.806	0.972	0.951
	✗	✗	499k	0.825	0.810	0.817	0.972	0.952
2D/3D CNN (ours)	✗	✗	321k	0.834	0.800	0.816	0.971	0.953
	✓	✗	413k	0.837	0.836	0.837	0.974	0.957
	✗	✓	321k	0.839	0.842	0.841	0.976	0.958
	✓	✓	414k	**0.897**	0.819	**0.856**	**0.978**	**0.964**

Table 4.2: Quantitative results of different deep learning model designs. The classification metrics are given in percent (%). Additionally, the total number of parameters (# Params) is provided. (TE) denotes the absolute temporal encoding and (LOAN) the location-aware adaptive normalization layer. The best and second-best results of each metric are highlighted in bold and italic text, respectively.

the LOAN layers increase the amount of parameters of the 2D/3D CNN, we report the results of the one-branch 3D CNN using 323k and 499k parameters. The smaller 3D CNN has about the same amount of parameters as the 2D/3D CNN without LOAN, whereas the larger 3D CNN has more parameters than the proposed model. Due to the lack of a spatial modeling, LSTM has the fewest parameters and is the fastest, but the precision is very low. ConvLSTM, the small one-branch 3D CNN, and 2D/3D CNN without LOAN and TE, which consider the spatial context, perform similar but 2D/3D CNN is the fastest approach and ConvLSTM is the slowest approach among them. Compared to the other approaches, transformer models have considerably more parameters and require more operations, which makes them computationally expensive. Nevertheless, our approach outperforms both transformer models while requiring by far less parameters and computational operations. Normalizing the dynamic features conditioned on the static features (LOAN) increases all metrics. It also outperforms the large 3D CNN in all metrics and inference time. Adding TE increases all metrics when LOAN is not used, while increasing the computational cost only very little. When LOAN is used, adding TE decreases the recall but increases all other metrics. Since LOAN and TE change the dynamic features, we observe a different trade-off between recall and precision if both are used. This change is consistent over the years as shown in Table 4.1. Nevertheless, the F1-Score, AUROC, and OA are highest if both are used.

State-of-the-art with LOAN. After the year 2023, a series of works were built upon the LOAN idea for wildfire forecasting, e.g., *Yin et al.* (2025) combined channel and spatial attention with LOAN and *Scudo et al.* (2025) adapted contrastive learning with LOAN for wildfire forecasting on FireCube

dataset. On a global scale, *Ji et al.* (2024) proposed a location-aware ConvLSTM and applied it for global wildfire forecasting.

Algorithm	LOAN	TE	Params(↓)	MMACs (↓)	# SPmS(↑)
LSTM (*Kondylatos et al.*, 2022)	✗	✗	30k	0.27	955±43
ConvLSTM (*Kondylatos et al.*, 2022)	✗	✗	372k	417.08	7±0
TimeSformer (*Bertasius et al.*, 2021)	✗	✗	1.16m	831,667.58	2±0
SwinTransformer (*Liu et al.*, 2022b)	✗	✗	1.78m	122,346.56	1±0
3D CNN	✗	✗	323k	476.56	18±1
	✗	✗	499k	585.11	17±1
2D/3D CNN (ours)	✗	✗	321k	137.24	47±0
	✓	✗	413k	168.23	33±1
	✗	✓	321k	137.24	47±1
	✓	✓	414k	168.23	33±2

Table 4.3: Comparison between different deep learning model designs. The total number of parameters (# Params), estimated multiply-accumulate operations given in Mega (MMACs), and the inference time as samples per millisecond (# SPmS) are provided. (TE) denotes the absolute temporal encoding and (LOAN) the location-aware adaptive normalization layer.

4.3.3 Qualitative Results

Predicted wildfire danger-susceptibility maps are depicted in Figs. 4.6-4.10. We take input from the 1^{st} and 2^{nd} days of three consecutive months in summer (June, July, and August) and predict for the 2^{nd} and 3^{rd} days of each month. We end up with around 500k pixels (samples) per day. The output from the deep learning models (LSTM, ConvLSTM, 3D CNN, TimeSFormer, SwinTransformer, Ours) is a probability $Y \in [0, 1]$. In addition, we visualize the predicted maps produced by FWI with the provided spatial resolution 8 km $\times$ 8 km. The output of FWI is clipped to the range $[0, 50]$ (*Kondylatos et al.*, 2022). The ignition points of large wildfires at those days are represented as black circles on the map. The first observation is that regardless of the coarse resolution of FWI, the predicted maps produced by the deep learning models are more reliable. While FWI relies on meteorological observations and models functional relationships, the results show that the modeled functional relationships are insufficient and do not reflect the complexity of the problem of forecasting wildfire. We also find that our proposed model with TE discards spots that result in a high false positive error rate while it keeps extreme ones (cf. the results for July in 2020 and 2021). Another important observation is that the LSTM model, which does not account for the spatial context, tends to produce heterogeneous predictions where neighboring pixels often have very different wildfire danger probabilities. Consequently, it generates many false positives. In contrast, our proposed model produces more homogeneous and clustered predictions. The respective performance metrics for Figs. 4.7-4.10 are provided in Table 4.4. As shown in Table 4.4, all models perform very well

in June 2020 and 2021. In these months, there are no positive samples and our model generates the least number of false positives. In July 2020 and 2021, only 2D/3D CNN predicts all positive samples whereas 2D/3D CNN with absolute temporal encoding generates the least false positives. The results are mixed in August. While TimeSformer performs very well in August 2020, our approach performs best in August 2021.

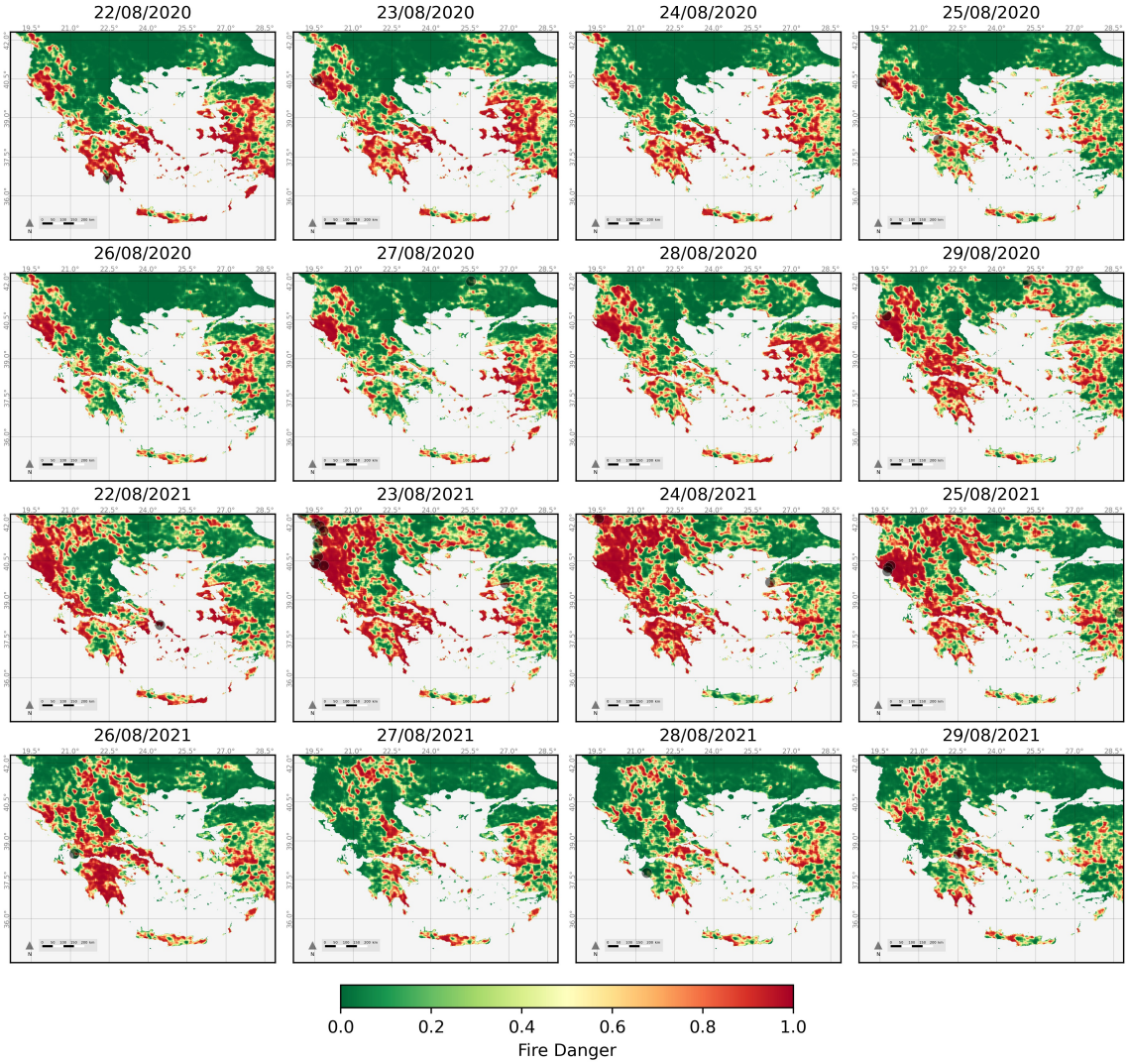


Figure 4.6: Predictions for 8 consecutive days in August. The upper two rows show the results for the year 2020 and the last two rows show the results for the extreme year 2021. The maps are produced by the proposed 2D/3D CNN with absolute temporal encoding. The black circles represent an ignition of a large wildfire on that day.

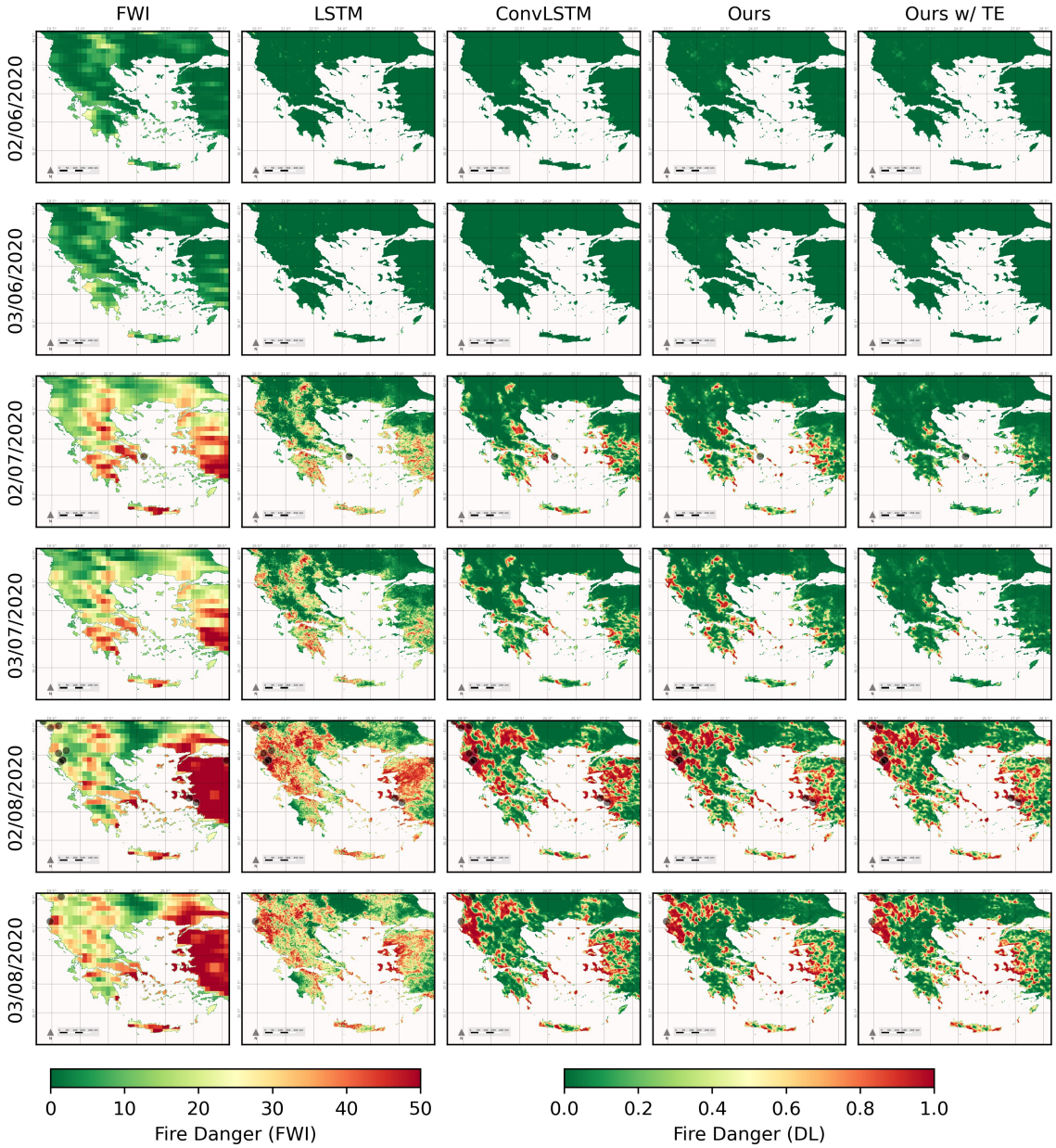


Figure 4.7: Qualitative results for 6 days during the wildfire season in year 2020. The black circles represent an ignition of a large wildfire on that day. (TE) denotes absolute temporal encoding.

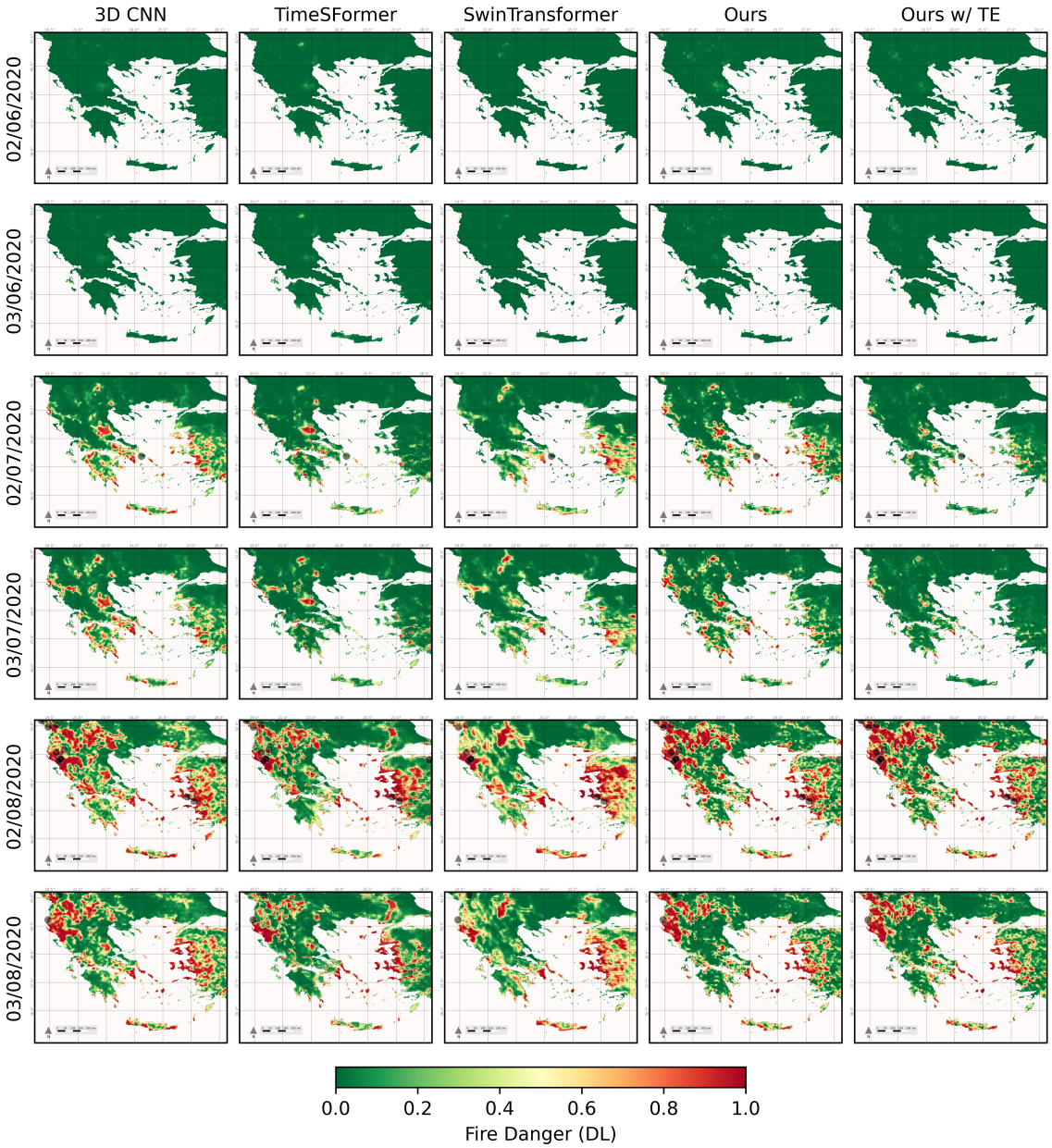


Figure 4.8: Qualitative results for 6 days during the wildfire season in year 2020. The black circles represent an ignition of a large wildfire on that day. (TE) denotes absolute temporal encoding.

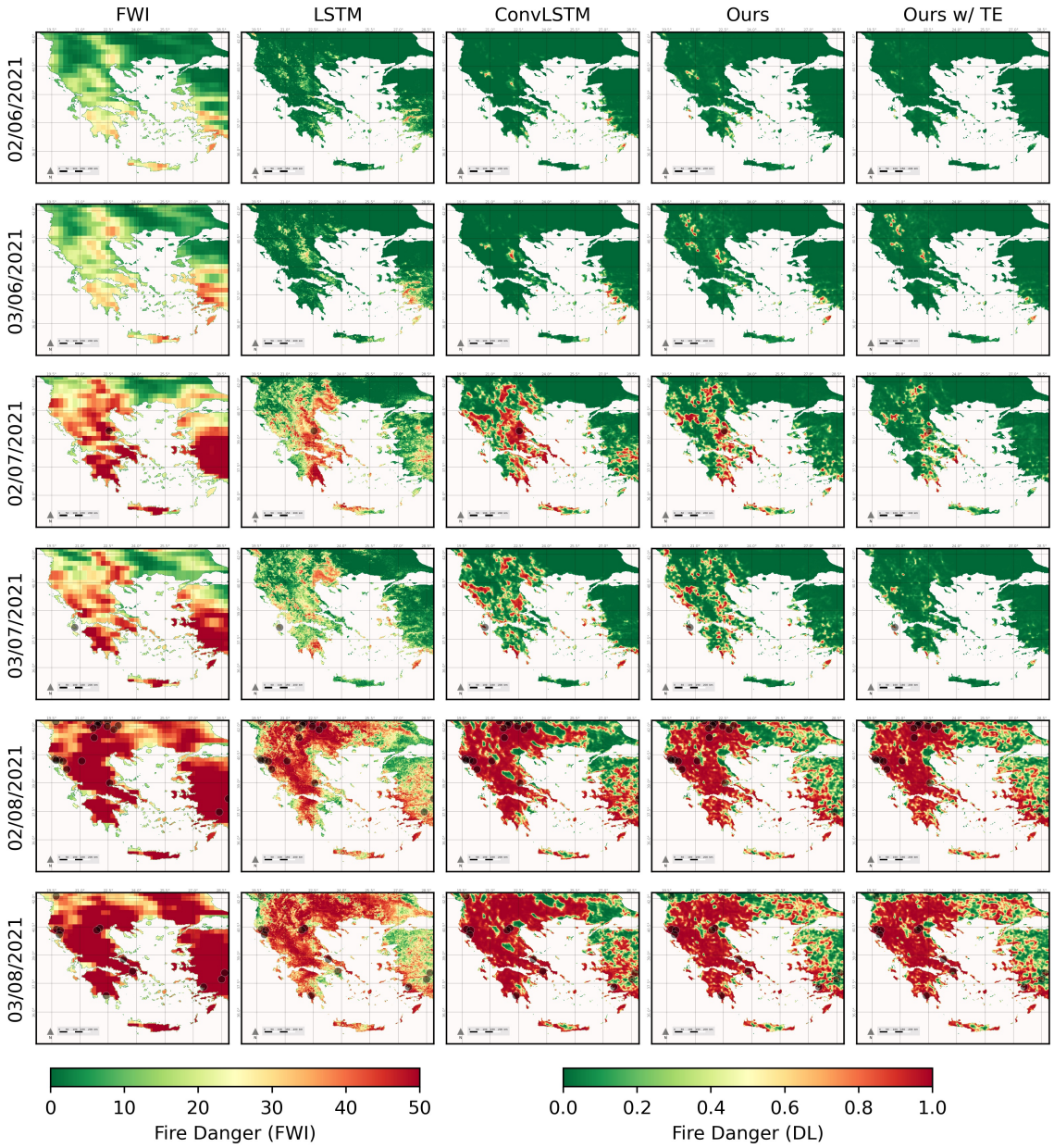


Figure 4.9: Qualitative results for 6 days during the wildfire season in the extreme year 2021. The black circles represent an ignition of a large wildfire on that day. (TE) denotes absolute temporal encoding.

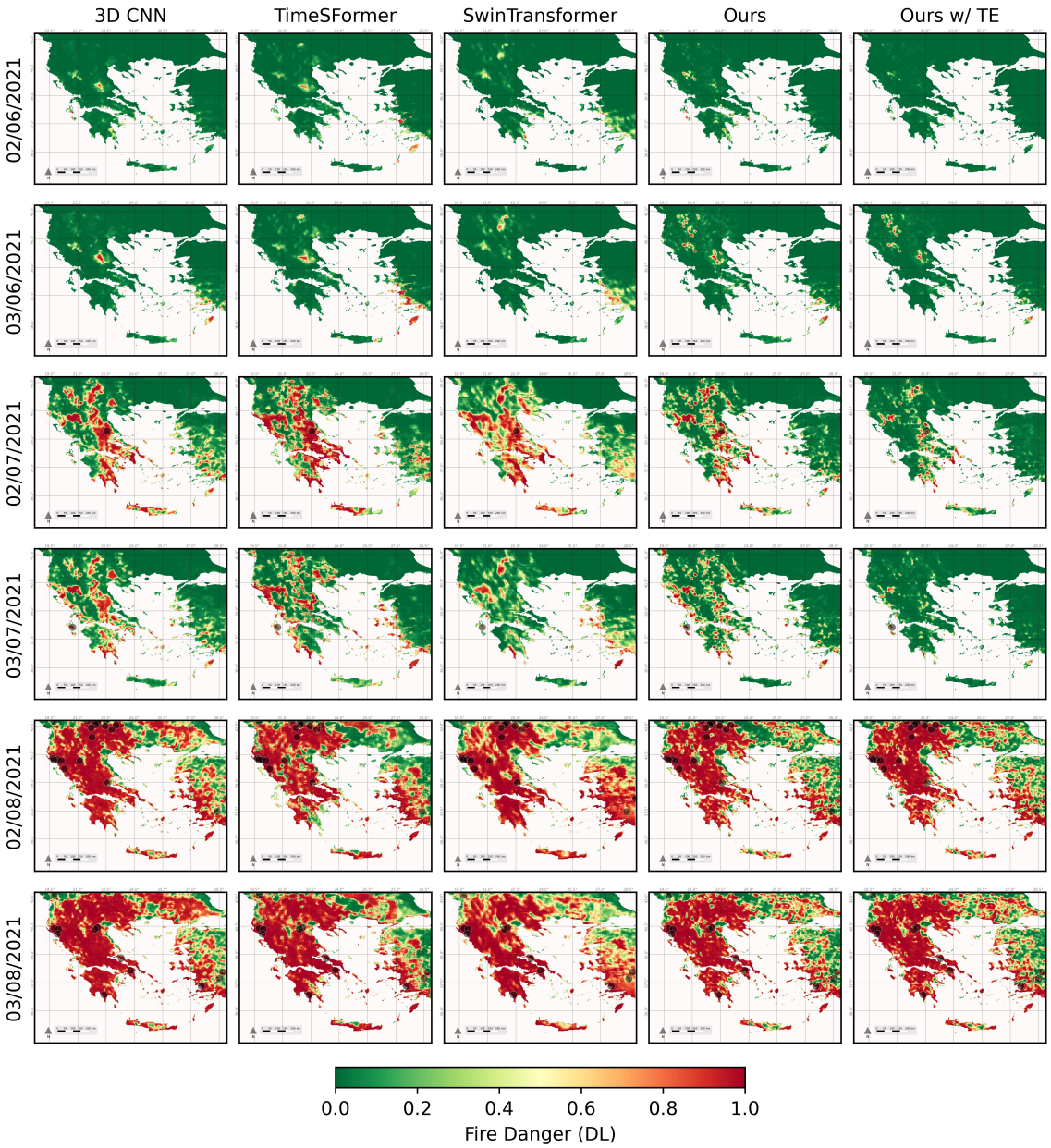


Figure 4.10: Qualitative results for 6 days during the wildfire season in the extreme year 2021. The black circles represent an ignition of a large wildfire on that day. (TE) denotes absolute temporal encoding.

# Positive samples	02/06/2020 0		03/06/2020 0		02/07/2020 2	
Algorithm	P	N	P	N	P	N
LSTM (*Kondylatos et al.*, 2022)	-	0.9998	-	0.9996	0.0000	0.8141
ConvLSTM (*Kondylatos et al.*, 2022)	-	**1.0000**	-	**1.0000**	0.0000	0.8972
TimeSformer (*Bertasius et al.*, 2021)	-	**1.0000**	-	**1.0000**	0.0000	0.9560
SwinTransformer (*Liu et al.*, 2022b)	-	**1.0000**	-	**1.0000**	0.0000	0.9119
3D CNN	-	**1.0000**	-	**1.0000**	0.0000	0.8828
2D/3D CNN (ours)	-	**1.0000**	-	0.9999	**1.0000**	0.8951
2D/3D CNN w/ TE (ours)	-	**1.0000**	-	**1.0000**	0.0000	**0.9681**

# Positive samples	03/07/2020 0		02/08/2020 71		03/08/2020 25	
Algorithm	P	N	P	N	P	N
LSTM (*Kondylatos et al.*, 2022)	-	0.8237	0.8732	0.4893	**0.8800**	0.5623
ConvLSTM (*Kondylatos et al.*, 2022)	-	0.9246	**0.9296**	0.6509	0.8000	0.7252
TimeSformer (*Bertasius et al.*, 2021)	-	0.9541	0.8310	**0.7102**	0.8000	**0.7824**
SwinTransformer (*Liu et al.*, 2022b)	-	0.9264	0.8873	0.6557	0.8661	0.7511
3D CNN	-	0.8744	0.8028	0.6383	0.8000	0.7106
2D/3D CNN (ours)	-	0.9033	0.7324	0.6674	0.8000	0.7617
2D/3D CNN w/ TE (ours)	-	**0.9792**	0.7887	0.6568	0.8000	0.7432

# Positive samples	02/06/2021 0		03/06/2021 0		02/07/2021 36	
Algorithm	P	N	P	N	P	N
LSTM (*Kondylatos et al.*, 2022)	-	0.9823	-	0.9644	**1.0000**	0.7058
ConvLSTM (*Kondylatos et al.*, 2022)	-	0.9911	-	0.9825	**1.0000**	0.7553
TimeSformer (*Bertasius et al.*, 2021)	-	0.9892	-	0.9738	**1.0000**	0.7346
SwinTransformer (*Liu et al.*, 2022b)	-	0.9929	-	0.9826	**1.0000**	0.7214
3D CNN	-	0.9952	-	0.9853	**1.0000**	0.7839
2D/3D CNN (ours)	-	0.9938	-	0.9808	**1.0000**	0.8436
2D/3D CNN w/ TE (ours)	-	**0.9991**	-	**0.9895**	**1.0000**	**0.9477**

# Positive samples	03/07/2021 22		02/08/2021 679		03/08/2021 1417	
Algorithm	P	N	P	N	P	N
LSTM (*Kondylatos et al.*, 2022)	0.6818	0.8314	0.8203	0.3254	0.9330	0.2388
ConvLSTM (*Kondylatos et al.*, 2022)	**1.0000**	0.8742	0.7187	0.3566	0.8701	0.2828
TimeSformer (*Bertasius et al.*, 2021)	0.9545	0.8184	0.4772	0.3886	0.9704	0.2899
SwinTransformer (*Liu et al.*, 2022b)	0.7727	0.9125	**0.9705**	0.3013	0.9471	0.2623
3D CNN	**1.0000**	0.8505	0.8351	0.2980	0.9259	0.2245
2D/3D CNN (ours)	**1.0000**	0.8897	0.8380	**0.3917**	0.9929	**0.3375**
2D/3D CNN w/ TE (ours)	**1.0000**	**0.9861**	0.8513	0.3667	**0.9936**	0.3167

Table 4.4: Recall (%) metric per class for Figs. 4.7-4.10. (P) denotes the positive class and (N) the negative one. The number of positive samples is shown below the date. This is equivalent to the area (km^2) that was burned by large wildfires at that day. The best results of each metric are highlighted in bold text.

Year 2019 (val)					
Static Variables	Precision(↑)	Recall(↑)	F1(↑)	AUROC(↑)	OA(↑)
DEM + slope	0.756	0.737	0.746	0.938	0.899
Distance to roads Distance to waterway Population density	0.773	0.761	0.767	0.942	0.906
Land cover	0.773	0.766	0.769	0.946	0.907
DEM + slope Distance to roads Distance to waterway Population density	0.751	**0.770**	0.760	0.941	0.902
DEM + slope Land cover	0.746	0.735	0.740	0.942	0.896
Land cover Distance to roads Distance to waterway Population density	0.775	0.732	0.753	**0.945**	0.903
All static variables	**0.796**	0.746	**0.771**	**0.945**	**0.910**
Year 2020-2021 (test)					
Static Variables	Precision(↑)	Recall(↑)	F1(↑)	AUROC(↑)	OA(↑)
DEM + slope	0.809	0.817	0.813	0.971	0.951
Distance to roads Distance to waterway Population density	0.819	0.745	0.780	0.961	0.945
Land cover	0.819	0.799	0.809	0.969	0.950
DEM + slope Distance to roads Distance to waterway Population density	0.796	0.769	0.782	0.967	0.944
DEM + slope Land cover	0.801	0.814	0.807	0.972	0.949
Land cover Distance to roads Distance to waterway Population density	0.826	0.811	0.819	0.970	0.953
All static variables	**0.837**	**0.836**	**0.837**	**0.974**	**0.957**

Table 4.5: Performance metrics for different categories of static variables. The values are given in percent (%). The best results of each metric are highlighted in bold text.

4.3.4 Variable Importance

To assess the importance of different static variables, we present the results obtained with different combinations of static variables in Table 4.5. For this experiment, we use 2D/3D CNN with LOAN but without TE. All dynamic variables are used in this experiment and the results are reported for the year 2019 and for the years 2020-2021 as one set. The static variables are grouped into 3 main categories: topographic variables consisting of digital elevation model (DEM) and slope, anthropogenic-related variables consisting of distance to roads or waterway and population density, and land cover variables.

From the results in Table 4.5, we can conclude that among the 3 categories topographic variables give the best results for the years 2020-2021 when they are used without other variables. While land cover and anthropogenic-related variables provide the best results for the year 2019. Overall, all static variables are relevant and the best results are obtained when all static variables are used (last row). This is also better than using the static variables of 2 out of the 3 categories, which is reported in rows 4-6 of Table 4.5.

4.3.5 Comparing Different Conditional Maps

While Table 4.5 shows the importance of different static variables as input to the 2D/3D CNN, we also analyze the impact of different ways to modulate the dynamic features in Table 4.6. For the experiments, we use the 2D/3D CNN without TE and all dynamic and static variables as input. While we use all variables, the different settings differ in the input that is fed to the LOAN layer, e.g., the static features that the modulation of the dynamic features is conditioned on.

The results of the proposed conditioning, where we use the features from the corresponding block of the static branch, are shown in the last row. In the first row, we show the results if we do not use LOAN at all, e.g., we do not use any modulation of the dynamic features. For the other rows of Table 4.6, we modify LOAN such that it is not conditioned on the intermediate features of the static branch as shown in Figs. 4.1 and 4.2. Instead, we condition LOAN directly on static variables. Note that we need to slightly adapt LOAN as shown in Fig. 4.3 since the number of static input variables C differs from the number of feature channels K at the block where LOAN is added.

As seen from Table 4.6, the best result is obtained when we use the activation maps, e.g., the intermediate features, from the static branch for conditioning the modulation. However, comparable results are obtained when all static variables are directly used by the variant of LOAN that is shown in Fig. 4.4. While this variant achieves slightly higher AUROC, the variant shown in Fig. 4.3 achieves higher F1-score and OA. Using the static variables directly, we can analyze how the three categories of static variables impact the modulation of the dynamic features and thus the results. We can conclude that the modulation of the dynamic features is very sensitive to its condition. For the year 2019, all three categories (rows 2-4) improve the results compared to the setup without feature modulation (first row). For the years 2020-2021, this is not the case and only the combination of all static variables (row 5) leads to an improvement. The reason is the mismatch between C and K^i. Compared to the other categories, land cover has the highest number of variables per pixels (C=10) and shows the best performance. This indicates that conditioning the modulation of the dynamic features on the intermediate features of the static branch is a more practical approach than conditioning the modulation on the static variables directly, which seems to be sensitive to the number of variables.

Year 2019 (val)					
Conditional Map	Precision(↑)	Recall(↑)	F1(↑)	AUROC(↑)	OA(↑)
–	0.757	0.622	0.683	0.934	0.883
DEM + slope	0.708	0.710	0.709	0.928	0.882
Distance to roads Distance to waterway Population density	0.738	**0.762**	0.750	0.937	0.897
Land cover	0.791	0.720	0.754	0.946	0.905
All static variables	0.771	0.749	0.760	**0.950**	0.904
Activation maps (Static branch)	**0.796**	0.746	**0.771**	0.945	**0.910**
Year 2020-2021 (test)					
Conditional Map	Precision(↑)	Recall(↑)	F1(↑)	AUROC(↑)	OA(↑)
–	0.834	0.800	0.816	0.971	0.953
DEM + slope	0.769	0.795	0.782	0.962	0.942
Distance to roads Distance to waterway Population density	0.780	0.766	0.773	0.962	0.941
Land cover	0.815	0.752	0.782	0.968	0.945
All static variables	0.811	**0.841**	0.825	**0.976**	0.953
Activation maps (Static branch)	**0.837**	0.836	**0.837**	0.974	**0.957**

Table 4.6: Impact of different conditional maps on the feature modulation. The evaluation metrics are given in percent (%). The best results of each metric are highlighted in bold text.

4.3.6 Ablation Studies

We finally evaluate two additional aspects. In Section 4.3.6.1, we analyze at which blocks of the proposed 2D/3D CNN LOAN is best added and the impact of the absolute temporal encoding with respect to the number of negative samples in Section 4.3.6.2.

4.3.6.1 LOAN Position in the Model

As shown in Fig. 4.1, the proposed network has three blocks and we add LOAN to the first and second block. We evaluate in Table 4.7 different configurations where we add LOAN only to the first or to all three blocks. The results are reported without TE. If we add LOAN only to the first block, the performance increases for the year 2019 but not for the years 2020-2021 compared to our model without LOAN (first row). When adding LOAN to the first two blocks, we observe a consistent improvement for all years. For the year 2019, Precision, Recall, F1-score, and AUROC are improved by $+3.96\%$, $+12.39\%$, $+8.75\%$, and $+1.14\%$, respectively, and by $+0.36\%$, $+3.62\%$, $+2.02\%$, and $+0.30\%$ for the years 2020-2021, respectively. Adding LOAN to all three blocks performs worse

Year 2019 (val)						
Block			Precision(↑)	Recall(↑)	F1-score(↑)	AUROC(↑)
1st	2nd	3rd				
✗	✗	✗	0.757	0.622	0.683	0.934
✓	✗	✗	0.769	**0.755**	0.762	0.944
✓	✓	✗	**0.796**	0.746	**0.771**	**0.945**
✓	✓	✓	0.739	0.709	0.724	0.937
Year 2020-2021 (test)						
Block			Precision(↑)	Recall(↑)	F1-score(↑)	AUROC(↑)
1st	2nd	3rd				
✗	✗	✗	0.834	0.800	0.816	0.971
✓	✗	✗	0.809	0.787	0.798	0.972
✓	✓	✗	**0.837**	**0.836**	**0.837**	**0.974**
✓	✓	✓	0.798	0.813	0.805	0.969

Table 4.7: Ablation study of different position choices for the LOAN layer. All classification metrics are given in percent (%).

than adding LOAN only to the first two blocks. This is due to the decrease of spatial resolution after each block by the pooling layers. In the third block the spatial resolution is too coarse for a location-specific modulation of the dynamic features.

4.3.6.2 Absolute Temporal Encoding

As we have seen in Tables 4.1 and 4.2, absolute temporal encoding (TE) increases precision at the cost of lower recall. Depending on the use of the wildfire forecasting, recall or precision are more important. The precision also depends on the amount of negative samples. In order to show that TE gives consistently a higher precision, we varied the number of negative samples. In this experiment, we test on all positive samples in the test set (years 2020-2021) and gradually increase the number of negative ones. As shown in Fig. 4.11, we start with a setup where the number of negative samples is equal to the number of positive samples, e.g., 5635 positive and 5635 negative samples. We then increase the number of negative samples. Since the number of negative samples increases, the precision decreases for all methods. Note that the recall does not change since the number of positive samples remains the same. As already observed in Table 4.2, LSTM has a very low precision. 2D/3D CNN with LOAN has in all settings a higher precision than ConvLSTM and a much higher recall as shown in Table 4.2. Transformer models on the other hand have less precision than ConvLSTM but provide an overall better recall as shown in Table 4.2. While adding TE decreases recall, it improves the precision substantially and the improvement increases when the number of negative samples increases. While other metrics like F1-score or AUROC combine precision and recall in a single measure, depending on the application a higher recall or a higher precision might be more important.

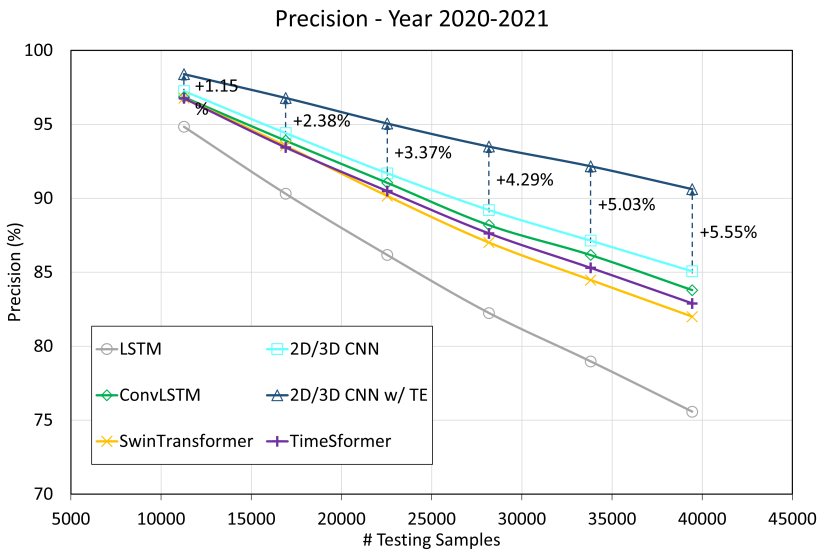


Figure 4.11: Impact of the number of testing samples on the precision. (TE) denotes the absolute temporal encoding. Here precision is multiplied by 100.

If precision is more important, TE is very useful. If recall is more important, TE should not be used. We also point out that TE encodes only the day of the year since the dataset has a relatively small spatial extension (10.2° Lon $\times$ 8° Lat). In case of larger datasets at continental scale, a consideration of the spatial location for the encoding would also become relevant as biogeographical regions occur (*Nyborg et al.*, 2022b), which are characterized by different climate variabilities and anthropogenic drivers over time.

Finally, we plot in Fig. 4.12 the loss (4.4) curve during training.

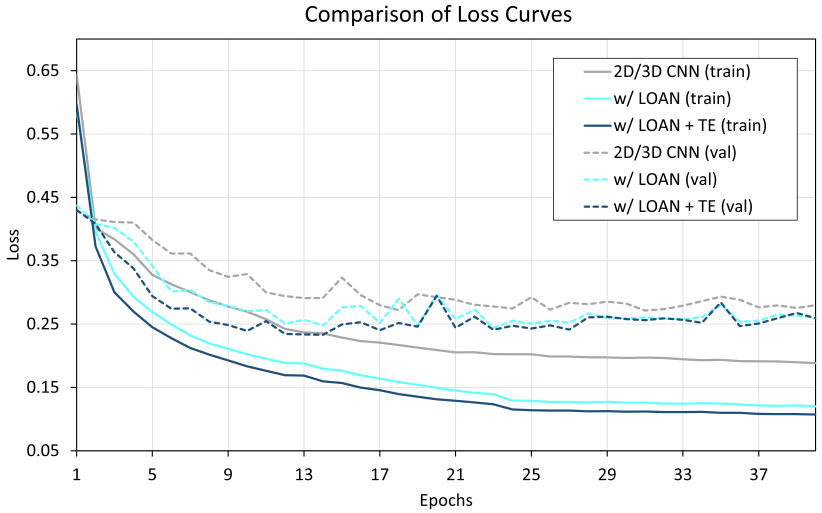


Figure 4.12: BCE loss during training for the training (solid) and validation (dashed) set. (LOAN) denotes the Location-aware Adaptive Normalization layer and (TE) denotes the absolute temporal encoding.

4.4 Summary

In this chapter, we proposed a new deep learning approach for wildfire danger forecasting. In contrast to previous works, we handle spatial (static) and spatio-temporal (dynamic) variables differently. Our model processes the spatial and spatio-temporal variables in two separated 2D/3D CNN branches to learn static and dynamic feature vectors. Moreover, we have introduced the Location-aware Adaptive Normalization layer, which modulates the activation maps in the dynamic branch conditionally on their respective static features to address the causal effect of static features on dynamic features. We furthermore integrated an absolute time encoding into the model. By encoding the calendar time, we make the model explicitly aware of the forecasting day. While the time encoding reduces the recall, it substantially increases the precision. We conducted our experiments on the FireCube dataset and demonstrated the effectiveness of our approach compared to several baselines in terms of Precision, F1-score, AUROC, and OA. We believe that the proposed approach of dealing with spatial and spatio-temporal variables is also highly relevant for other remote sensing applications.

CHAPTER 5

Focal-TSMP: Deep Learning for Vegetation Health Prediction and Agricultural Drought Assessment from a Regional Climate Simulation

---

This chapter is based on the following publication (*Shams Eddin and Gall*, 2024a):

Focal-TSMP: Deep Learning for Vegetation Health Prediction and Agricultural Drought Assessment from a Regional Climate Simulation
Mohamad Hakam Shams Eddin, and Juergen Gall
Geoscientific Model Development (GMD), 2024
Doi: 10.5194/gmd-17-2987-2024
Url: https://hakamshams.github.io/Focal-TSMP/

Below we summarize the contributions of each author.

- **Mohamad Hakam Shams Eddin**
  Hakam proposed the idea of using focal modulation and designed the method to predict agricultural drought. Hakam also proposed to use the remote sensing indices. He designed the experimental setup, conducted the experiments, preprocessed the data, implemented the codebase, and wrote the paper.

- **Juergen Gall**
  Juergen proposed the idea of predicting drought with a TSMP simulation. He supervised and funded the project. In addition, he contributed with discussions and writing.

Contents

5.1 Introduction 79
5.2 Methodology 81
5.2.1 Model Architectures 81
5.2.2 Focal Modulations 83
5.2.3 Channel Attention 86
5.2.4 Loss Function 87
5.3 Experimental results and analysis 87
5.3.1 Experimental Setup 88

5.3.2 NDVI and BT Prediction 89
5.3.3 Agricultural Drought Assessment 98
5.3.4 Variable Importance 101
5.3.5 Ablation Studies 104
5.4 Summary 106

In the previous chapter, we addressed the task of wildfire forecasting at a local scale. In this chapter, we address the task of forecasting impact of extreme events at a continental scale. We consider agricultural drought as an example of impacts of extremes on vegetation. To define agricultural droughts, we rely on satellite-derived indices. These agricultural drought indices can provide a complementary perspective of terrestrial vegetation trends. In addition, their integration for drought assessments under future climates is beneficial for providing more comprehensive assessments. However, satellite-derived drought indices are only available for the Earth observation era. In this chapter, we aim thus to improve the agricultural drought assessments under future climate change by applying deep learning (DL) to predict satellite-derived vegetation indices from a regional climate simulation. The simulation is produced by the Terrestrial Systems Modelling Platform (TSMP) and performed in a free evolution mode over Europe. TSMP simulations incorporate variables from underground to the top of the atmosphere (Ground to Atmosphere G2A) and are widely used for research studies related to water cycle and climate change. We leverage these simulations for long-term forecasting and DL to map the forecast variables into Normalized Difference Vegetation Index (NDVI) and Brightness Temperature (BT) images that are not part of the simulation model. These predicted images are then used to derive different vegetation and agricultural drought indices, namely NDVI anomaly, BT anomaly, Vegetation Condition Index (VCI), Thermal Condition Index (TCI), and Vegetation Health Index (VHI). The developed DL model in this chapter could be integrated with data assimilation and used for down-stream tasks, e.g., for estimating NDVI and BT for periods where no satellite data are available and for modelling the impact of extreme events on vegetation responses with different climate change scenarios. Moreover, the model could be used as a complimentary evaluation framework for TSMP-based climate change simulations. To ensure reliability and to assess the model's applicability to different seasons and regions, we provide in this chapter an analysis of model biases and uncertainties across different regions over the Pan-Europe domain. We further provide an analysis about the contribution of the input variables from the TSMP model components to ensure a better understanding of the model prediction. A comprehensive evaluation on the long-term TSMP simulation using reference remote sensing data showed sufficiently good agreements between the model predictions and observations. While model performance varies on the test set between different climate regions, it achieves a mean absolute error (MAE) of 0.027 and 1.90 K with coefficient of determination (R^2) scores of 0.88 and 0.92 for NDVI and BT, respectively, at 0.11° resolution for sub-seasonal predictions. In summary, we demonstrate in this chapter the feasibility of using DL on a TSMP simulation to synthesize NDVI and BT satellite images, which can be used for agricultural drought forecasting.

5.1 Introduction

According to recent studies on historical trends and current projections, different regions of the Earth would be under a changing climate more vulnerable to extreme events such as flash droughts (*Christian et al.*, 2021, 2023; *Yuan et al.*, 2023), meteorological and agricultural droughts (*Essa et al.*, 2023), forest wildfires (*Patacca et al.*, 2023), and water storage deficiency (*Pokhrel et al.*, 2021). The expected increase in concurrence of agricultural droughts would cause crop production losses and vegetation mortality. In particular, people in regions with fragile adaptation and mitigation strategies will be more effected. Therefore, forecasting the vegetation responses and their evolving patterns conditioned on climate scenarios is a requirement to form better mitigation and adaptation strategies.

In relation to this, there has been a growing line of research over the past in improving and deploying climate modeling that attempt to simulate the underlying processes of the Earth system (*Shrestha et al.*, 2014; *Gasper et al.*, 2014; *Lawrence et al.*, 2019). These modelling platforms are essential to realize and forecast climatic extreme events such as droughts in a model simulation (*Miralles et al.*, 2019). For instance, the simulated outputs of modeling systems can be used to derive agricultural drought indices based on a deficiency in precipitation (*McKee*, 1995; *Vicente-Serrano et al.*, 2010) or soil moisture (*Martínez-Fernández et al.*, 2015). Nowadays, satellite observations around the world provide a near real-time global monitoring of vegetation and drought conditions. Vegetation products derived from satellite land surface reflectances can be used as proxies for vegetation health and consequently as agricultural drought indicators (*Qin et al.*, 2021; *Vreugdenhil et al.*, 2022). While historical trends in satellite-based droughts have been extensively studied, satellite-based agricultural drought assessment and its relation to climate simulations under climate change remains not fully explored.

In this chapter, we propose to use deep learning (DL) to improve the agricultural drought analysis by predicting satellite-derived vegetation indices that can be combined with meteorological or hydrological indices which are often used in studies for drought assessment to provide more comprehensive assessments. In fact, some studies highlighted inconsistencies in the long-term drought trends (*Sheffield et al.*, 2012; *Kew et al.*, 2021; *Vicente-Serrano et al.*, 2022). Meanwhile others showed a different perspective of trends related to terrestrial vegetation from remote sensing products (*Zhu et al.*, 2016; *Kogan et al.*, 2020). This is usually explained as assessments are highly dependent on drought definition (*Satoh et al.*, 2021; *Reyniers et al.*, 2023) and extreme event attribution (*Van Oldenborgh et al.*, 2021), e.g., the drought indicator that was chosen in the methodology and the variations in modelling platforms. In addition, prescribed vegetation assumptions exist in climate simulations which limit the modeling of atmospheric carbon effects or soil moisture deficiency on vegetation (*Pirret et al.*, 2020; *Pokhrel et al.*, 2021; *Reyniers et al.*, 2023). If we add to this the complex spatio-temporal response of vegetation to climate variability (*Seneviratne et al.*, 2021; *Jin et al.*, 2023), e.g., regional responses to climate have different dynamics and are more complicated than those at a global scale, we can conclude that predicting the vegetation state in response to drought under climate conditions still poses a major challenge. More precisely, in this chapter, we predict satellite-based vegetation products from a free evolving simulation based on the Terrestrial Systems Modelling Platform (TSMP) (*Furusho-Percot et al.*, 2019). TSMP simulations integrate variables from groundwater to the top of the atmosphere (Ground to Atmosphere G2A) and are primarily employed in studies on the water cycle and climate change (*Ma et al.*, 2021; *Furusho-Percot et al.*, 2022; *Naz et al.*, 2023; *Patakchi Yousefi and Kollet*, 2023). Details about the TSMP simulation are provided

in Section 2.2.2.1. In particular, we predict from the TSMP simulation the Normalized Difference Vegetation Index (NDVI) and Brightness Temperature (BT) as they would have been observed from Advanced Very High Resolution Radiometer (AVHRR) from the National Oceanic and Atmospheric Administration (NOAA) satellite systems. NDVI is computed from the reflectance in visible red (ρ_R) and near-infrared bands (ρ_{NIR}). It is a standard product that is extensively used in applications for vegetation health and crop yield (*Tucker*, 1979). BT is a calibrated spectral radiation derived from the thermal band (ρ_{IR}) and can be used for temperature-related vegetation stress monitoring (*Kogan*, 1995b). We assume that a climate simulation (e.g., TSMP simulation) that is close to the true state of the Earth should be able to model vegetation products (e.g., NDVI and BT) regardless of the target satellite platform (in this chapter AVHRR NOAA). In this chapter, we thus propose a DL approach based on focal modulation networks (*Yang et al.*, 2022) to simultaneously predict NDVI and BT from the model simulation. In this way, we leverage a climate simulation for long-term forecasting and DL for mapping the forecast variables to vegetation related indices that are not part of the simulation model.

Forward operators like radiative transfer solvers are normally used to synthesize spectral band satellite images from the output of a numerical weather model (*Scheck et al.*, 2016; *Geiss et al.*, 2021; *Li et al.*, 2022a). In this chapter, we investigate the use of DL to predict products of atmospherically corrected albedo/emissivity on land (atmospherically corrected bottom of atmosphere) like NDVI and BT simultaneously rather than training the neural network to serve as an emulator for a predefined physical-based radiative transfer model. In other words, our training data for DL are derived from real-world satellite observations (empirical operator) without assimilating data or assumptions about radiations.

Unlike aforementioned works (Section 3.2.2), we use more input data for the neural network from the surface and sub-surface to account for a more detailed representation of the reflectance/emissivity on ground. In addition, we built the neural network on Vision Transformers (*Dosovitskiy et al.*, 2021) and Convolutional Neural Networks (CNN) models taking into account the spatial context around each input pixel and operating on the whole scene at once. This was motivated by previous studies that indicate that an effective model of the environment should consider the spatial-correlation within the domain. Previous works (Section 3.2.2) train and evaluate DL models on biased-corrected reanalysis data. In contrast, we evaluate the approach with real-world observations using a run of the simulation in the past. It is worth to note that this evaluation is more consistent with real-world deployment schemes, since it is questionable how a model that has been trained and evaluated on reanalysis data will perform on a biased climate projection simulations. Thus, we opt for a simulation that mimics a climate projection of the past and train and evaluate the model on it to internally correct biases and predict vegetation products. Moreover, as mentioned above, we aim to use DL to predict vegetation products like NDVI and BT from a regional climate simulation at a continental scale. We also focus on long-term forecasting without using an initial state, e.g., satellite images from previous time steps.

To showcase the potential of our approach, we apply the predicted NDVI and BT for long-term agricultural drought forecasting, where we derive Vegetation Condition Index (VCI), Thermal Condition Index (TCI), and Vegetation Health Index (VHI) (*Yang et al.*, 2020) as agricultural drought indicators from the predicted NDVI and BT (Section 2.2.3.2). As part of this, we analyze whether a DL model trained on simulation produced by TSMP can be used for vegetation health forecasting at a continental-scale by identifying regions and periods of uncertainty in the model prediction.

Moreover, we analyze the importance of the input explanatory variables. We achieve an overall mean absolute error (MAE) of 0.027 and 1.90 K with coefficient of determination (R2) scores of 0.88 and 0.92 in predicting NDVI and BT, respectively for sub-seasonal predictions at 0.11° resolution. Our results indicate that a direct prediction of vegetation products from a TSMP simulation with DL is an effective way for scenario-based assessments of vegetation response to climate change.

The rest of this chapter is organized as follows. The methodology is described in Section 5.2. Experimental results and analyses are given in Section 5.3. Finally, a summary is provided in Section 5.4.

5.2 Methodology

Problem formulation. Given TSMP $\in \mathbb{R}^{V\times T\times W\times H}$ as a climate change simulation, where V is the number of output variables from the COSMO, CLM, and ParFlow models and the static forcing variables, T is the temporal dimension and W and H are the spatial extensions, our objective is to construct a mapping function to predict NDVI $\in \mathbb{R}^{I\times W\times H}$ and BT $\in \mathbb{R}^{I\times W\times H}$ on a weekly basis, where I is the number of weeks. To accomplish this, we propose to approximate this function as a function f using a DL model based on a U-Net (*Ronneberger et al.*, 2015) with focal modulations (*Yang et al.*, 2022) as building blocks:

$$f : (\text{TSMP}; \theta) \rightarrow \ (\text{NDVI}, \text{BT}), \tag{5.1}$$

where θ are the weights of the model. The input for DL is a data cube representing a specific week i of TSMP data and the output are NDVI and BT corresponding to the same week i. We denote the weekly averaged input data cube produced by TSMP as $\mathbf{X}^i \in \mathbb{R}^{V\times W\times H}$. Where we obtain $\mathbf{X}^i$ by taking the mean of the days corresponding to the week i. For simplicity, we will drop the notation i in the following sections. First, the network architecture is introduced in Sect. 5.2.1 and the focal modulation is then described in Sect. 5.2.2. Section 5.2.3 discusses the channel attention for feature importance and Section 5.2.4 discusses the loss function.

5.2.1 Model Architectures

We build our model on focal modulation networks (Section 2.1.6). Figure 5.1 provides an overview of the model architecture. The model design follows the U-Net shape with encoder and decoder layers connected via skip connections and followed by two regression heads. This allows the model to extract features in a hierarchical way and predict NDVI and BT with customized heads. In the following, we describe the main parts of the model:

Patch embedding. The patch embedding is implemented as a single 1D convolution, where one patch is equivalent to one pixel. The role of this embedding is to project the input $\mathbf{X}$ from V dimension into a channel dimension that matches the channel dimension $C_{(en,1)}$ of the first encoder block. In contrast to related works with transformers, we do not reduce the spatial resolution at this step. This is important to mitigate blurring effects for regression tasks. An analysis of the impact of the patch size for embedding is provided in Section 5.3.5.2.

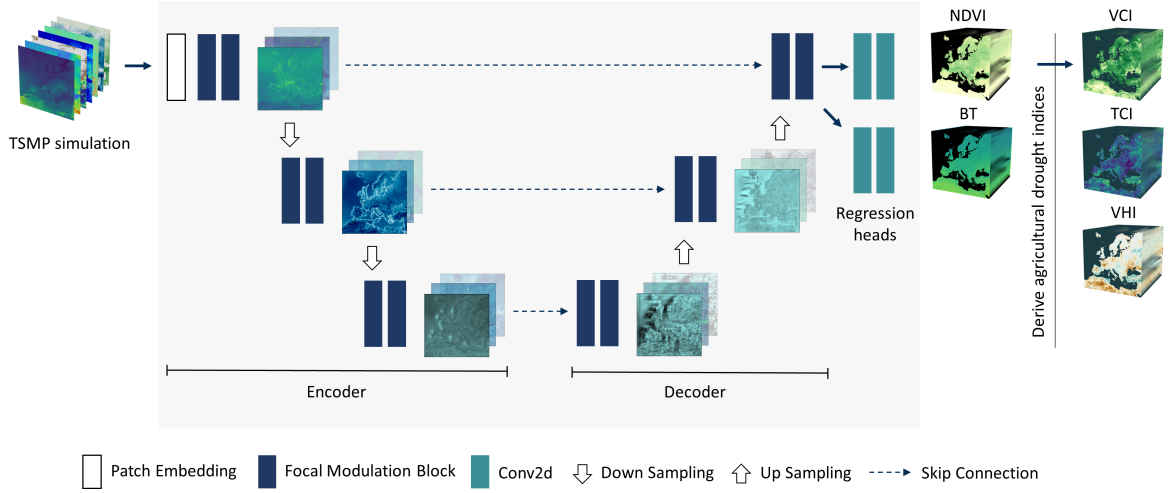


Figure 5.1: An overview of the proposed model to predict NDVI and BT from a TSMP climate simulation. The model follows the U-Net shape with encoder and decoder layers. We use focal modulation as the basic building block for the model. The input TSMP simulation is first encoded into a latent representation via encoder layers. In a subsequent step, the decoder constructs new features to be given as input to two separated regression heads that output NDVI and BT simultaneously. The predicted NDVI and BT can then be used to derive different agricultural drought indices such as VCI, TCI, and VHI.

Encoder. The encoder consists of 3 encoding layers. Each layer has 2 consecutive focal modulation blocks that have the same number of channels. We use focal modulation to capture local to global dependencies in the domain (Section 5.2.2). We apply down-sampling on the output of the first two encoder layers to reduce the spatial resolution by a factor of 2 and double the number of channels. The down-sampling is implemented as a 2D convolution with 2×2 kernel size and stride of 2. We set $C_{(en,1)} = 96$ as the number of channels of the first encoder layer. Consequently, the encoder has the dimensionality $\{C_{(en,1)} = 96, C_{(en,2)} = 192, C_{(en,3)} = 384\}$, where $C_{(en,2)}$ is the dimensionality for the second encoder layer and $C_{(en,3)}$ is the dimensionality for the third encoder layer. The encoder allows the network to extract low to high level features in a hierarchical way. Note that focal modulation allows an additional hierarchical feature extraction at each level (Section 5.2.2).

Skip connections. These connections copy outputs from each encoder layer into its corresponding decoder layer. The purpose of this is to enhance the gradient flow in the network and to mitigate vanishing gradient issues.

Decoder. The decoder has a similar design to the encoder. It consists of 3 decoder layers with 2 consecutive focal modulation blocks for each decoder layer. The input for the first decoder layer is the output of the last encoder layer copied via a skip connection. The input for the second and third decoder layers is a concatenation of the output from the previous decoder layer with the output of the corresponding encoder layer. The outputs of the first and second decoder layers are up-sampled to double the image size and reduce the dimensionality by a factor of 2. The up-sampling is imple-

mented as a bilinear interpolation followed by a 2D convolution with 1×1 kernel size and stride of 1. The decoder layers has the dimensionality $\{C_{(de,1)} = C_{(en,3)} = 384, C_{(de,2)} = C_{(en,2)} + C_{(de,1)} = 384, C_{(de,3)} = C_{(en,1)} + C_{(de,2)} = 288\}$, where $C_{(de,1)}$, $C_{(de,2)}$, and $C_{(de,3)}$ are the dimensionality for the first, second, and third decoder layers, respectively. The purpose of the decoder is to gradually construct the input for the regression heads from the encoded features.

Regression heads. The ouput of the last decoder layer is then given as input to two separated regression heads to predict NDVI and BT. Each head has two 2D convolutions with 3×3 kernel size and stride of 1 with a LeakyReLU activation in between. The regression head reduces the dimensionality from $C_{(de,3)} = 288$ to 128, and then to 1.

5.2.2 Focal Modulations

We first describe how the block is implemented in Section 5.2.2.1 and then describe the main focal modulation module denoted as FM in Section 5.2.2.2.

5.2.2.1 Focal Modulation Block

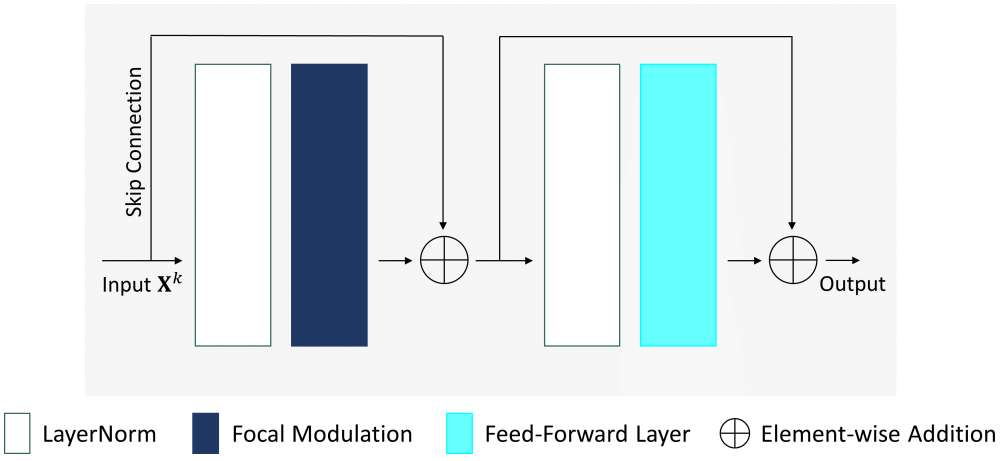


Figure 5.2: An illustration of the focal modulation block. It follows the typical transformer block with a focal modulation instead of self-attention. $\mathbf{X}^k$ represents the input to the k-th block.

Fig. 5.2 illustrates the architecture of the focal modulation block used in both the encoder and decoder layers. The design follows a typical transformer block. Let $\mathbf{X}^k \in \mathbb{R}^{N \times C^k \times W^k \times H^k}$ be the input at the k-th block, where N is the batch size (number of input tensors), C^k is the number of input channels, and W^k and H^k are the spatial resolution.

First, the input is normalized across N via a layer normalization (*Ba et al.*, 2016) denoted as LayerNorm. Using the indices $n \in \{1, \ldots, N\}$, $c^k \in \{1, \ldots, C^k\}$, $w^k \in \{1, \ldots, W^k\}$, and $h^k \in$

$\{1, \ldots, H^k\}$, the LayerNorm can be written as:

$$\text{LayerNorm}(\mathbf{X}^k; (\gamma_l^k, \beta_l^k)) = \left(\frac{\mathbf{X}^k_{n(c^k,w^k,h^k)} - \mu_n^k}{\sigma_n^k} \right) \cdot \gamma^k_{l\,(c^k)} + \beta^k_{l\,(c^k)}\,, \tag{5.2}$$

$$\mu_n^k = \frac{1}{C^k W^k H^k} \sum_{c^k=1}^{C^k} \sum_{w^k=1}^{W^k} \sum_{h^k=1}^{H^k} \mathbf{X}^k_{n(c^k,w^k,h^k)}\,, \tag{5.3}$$

$$\sigma_n^k = \sqrt{\frac{1}{C^k W^k H^k} \sum_{c^k=1}^{C^k} \sum_{w^k=1}^{W^k} \sum_{h^k=1}^{H^k} (\mathbf{X}^k_{n(c^k,w^k,h^k)} - \mu_n^k)^2}\,, \tag{5.4}$$

where $\mathbf{X}^k_{n(c^k,w^k,h^k)}$ is the input tensor of order n in the batch, μ_l^k and σ_l^k are the computed mean and standard deviation of the corresponding input $\mathbf{X}^k_{n(c^k,w^k,h^k)}$, and $\gamma^k_{l\,(c^k)} \in \mathbb{R}^{C^k}$ and $\beta^k_{l\,(c^k)} \in \mathbb{R}^{C^k}$ are per-element learnable parameters. These learnable parameters are shared across input tensors.

The output of LayerNorm is then passed into the function FM. After that, the output of the first part is normalized by a second LayerNorm and passed into a feed-forward layer (FFL). The FFL consists of one linear layer that maps the dimentionality to $r_{mlp} \times C^k$ followed by a GELU activation (*Hendrycks and Gimpel*, 2016) and a second linear layer to bring the dimensionality back to C^k, where r_{mlp} is the MLP ratio parameter. We set r_{mlp} to 4 for the encoder and decrease it to 2 for the decoder to reduce the number of model parameters. The output of each block can be formulated as follows:

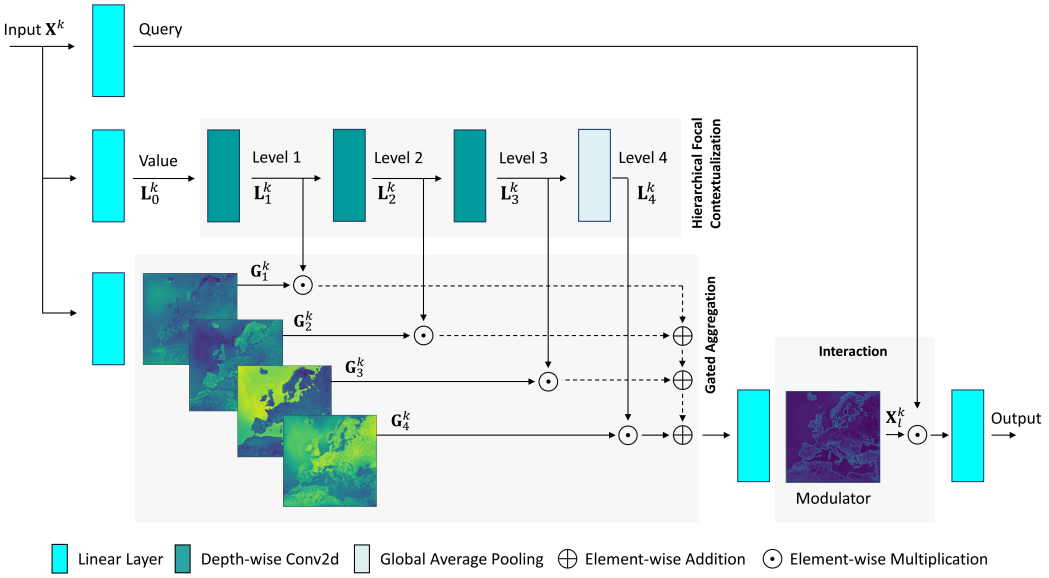
$$\begin{aligned}\text{FocalModulationBlock}(\mathbf{X}^k) \triangleq \boldsymbol{\zeta}_2^k \Big(\text{FFL}\big(\text{LayerNorm}_2(\boldsymbol{\zeta}_1^k \text{FM}(\text{LayerNorm}_1(\mathbf{X}^k)) + \mathbf{X}^k)\big) \Big) \\ + \Big(\boldsymbol{\zeta}_1^k \text{FM}(\text{LayerNorm}_1(\mathbf{X}^k)) + \mathbf{X}^k \Big)\,,\end{aligned} \tag{5.5}$$

where $\boldsymbol{\zeta}_1^k \in \mathbb{R}^{C^k}$ and $\boldsymbol{\zeta}_2^k \in \mathbb{R}^{C^k}$ are learnable scaling parameters.

5.2.2.2 Focal Modulation Module

The main component of each focal modulation block is the FM function. As seen in Fig. 5.3, it consists of three main steps: hierarchical contextualization, gated aggregation, and interactions.

Hierarchical contextualization. The objective of this part is to encode local to global range dependencies for every pixel. It is based on focal transformer (*Yang et al.*, 2021a) and aims to extract features at 4 different levels. Let $\mathbf{X}^k$ be the input for FM and $L = 4$ be the number of levels. First, $\mathbf{X}^k$ is projected by a linear layer into a new representation $\mathbf{L}_0^k = \text{Linear}(\mathbf{X}^k) \in \mathbb{R}^{N \times C^k \times W^k \times H^k}$. Afterwards, the contexts are obtained in a recursive manner using a sequence of 3 depth-wise 2D convolutions (DWConv2D) with GELU activation and with increased receptive fields. In DWConv2D, each output channel corresponds to a convolution on one input channel. We denote r_l as the kernel size at level l and start with $r_1 = 3$. Thereby, the kernel sizes at the focal levels have the values r_1=3, r_2=5, r_3=7. To obtain a global feature representation, a global average pooling (GAP) followed



.

Figure 5.3: An illustration of the function FM at k-th block. It consists of 3 main parts: focal contextualization, gated aggregation, and interaction. First, the query, value and gates are obtained by projecting $\mathbf{X}^k$ with linear layers. Then, a stack of depth-wise 2D convolutions followed by a global pooling is used on the value to derive contextual features around pixels. Gates are used to adaptive aggregate contextual features into a modulator. Finally, the interaction between queried pixels and the modulator is performed and projected by a final linear layer to compute the output. The shown images are examples of learned gates along with the pixel-wise magnitude of corresponding modulator at the first block encoder. The bright colors (e.g., green to yellow) for specific regions represent higher values which correspond to higher attentions of the model to that regions

by a GELU activation is applied at level $l = 4$. Using the index $l \in \{1, \ldots, L\}$, the hierarchical contextualization can be formulated as follows:

$$\mathbf{L}_l^k \triangleq \begin{cases} \text{GELU}\big(\text{DWConv2D}(\mathbf{L}_{l-1}^k)\big), & \text{if } 1 \leqslant l < L, \\ \text{GELU}\big(\text{GAP}(\mathbf{L}_{l-1}^k)\big), & \text{otherwise.} \end{cases} \tag{5.6}$$

Gated aggregation. The gated aggregation adaptively summarizes the extracted hierarchical contexts $\mathbf{L}_l^k$ into a modulator. First, $\mathbf{X}^k$ is projected by a linear layer into 4 gates $\mathbf{G}^k = \text{Linear}(\mathbf{X}^k) \in \mathbb{R}^{N \times L \times W^k \times H^k}$. As can be seen from the example in Fig. 5.3, the third gate focuses on the water area while other gates focus on different segmented regions. This allows each pixel to adaptively aggregate features from different semantic regions conditioned on its context. Pixels in a less dynamic environment may depend on more distant pixels while pixels in a more dynamic environment may depend more on the local context. The aggregation is performed over different focal levels and followed by a linear layer:

$$\mathbf{X}_L^k \triangleq \text{Linear}\Big(\sum_{l=1}^{L} \mathbf{G}_l^k \odot \mathbf{L}_l^k\Big), \tag{5.7}$$

where $\mathbf{X}_L^k \in \mathbb{R}^{N\times C^k\times W^k\times H^k}$ are the contextual aggregated features for each pixel called the modulator, $\mathbf{G}_l^k$ is the gate corresponding to level l, and $\odot$ is the Hadamard operator (element-wise multiplication).

Interaction. Finally, the interaction between the queried pixels and the modulator is given with the following formula:

$$\mathrm{FM}(\mathbf{X}^k) \triangleq \mathbf{X}_L^k \odot \mathrm{Linear}(\mathbf{X}^k) \in \mathbb{R}^{N\times C^k\times W^k\times H^k}. \tag{5.8}$$

5.2.3 Channel Attention

Channel attention (*Woo et al.*, 2018; *Hu et al.*, 2018) is commonly used in the field of computer vision and remote sensing to enhance feature representations inside DL models. In this chapter, we propose to use channel attention to determine the relative importance of TSMP input variables.

A channel attention module aims to calibrate the input variables/channels by learning an input-dependent scale for each channel. Thus, it can model the inter-correlation across variables adaptively (Fig. 5.4). In other words, channel attention aims to condense the input channels into a lower dimensionality and then it constructs channel scales with a sigmoid activation function ($\mathrm{Sigmoid}(x) = \frac{1}{1+e^{-x}} \in [0,1]$). In this manner, the neural network learns to calibrate the input channels with the learned scaling depending on the input channels. Given $\mathbf{X} \in \mathbb{R}^{V\times W\times H}$ as input TSMP simulation, where V is the number of output variables from COSMO, CLM, and ParFlow models, and W and H are the spatial extensions, the channel attention is computed as follows:

$$\mathrm{ChannelAttention}(\mathbf{X}) \triangleq \mathrm{Sigmoid}\Big(\mathrm{MLP}\big(\mathrm{GAP}(\mathbf{X})\big)+\mathrm{MLP}\big(\mathrm{GSD}(\mathbf{X})\big)\Big) \in \mathbb{R}^{V\times 1\times 1}, \tag{5.9}$$

where Sigmoid is the sigmoid function, MLP consists of two linear layers with a ReLU activation in between. The first decreases the dimension to $\frac{V}{r_{att}}$ and the subsequent layer maps it back to V. GAP is global average pooling, and GSD is the global standard deviation.

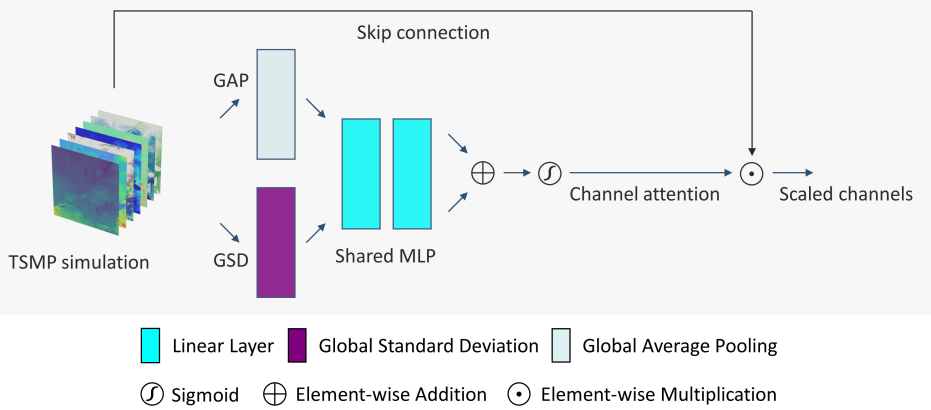


Figure 5.4: Illustration of the channel attention implementation. The output of channel attention is multiplied with the input TSMP to scale the channels from COSMO, CLM, and ParFlow according to their activation values.

5.2.4 Loss Function

For training we use the Mean Absolute Error (MAE) as a loss function, since it is less sensitive to outliers than the Mean Squared Error (MSE):

$$\mathcal{L}_{MAE} = \frac{1}{NWH} \sum_{n=1}^{N} \sum_{w=1}^{W} \sum_{h=1}^{H} |\mathbf{Y}_{(n,w,h)} - \hat{\mathbf{Y}}_{(n,w,h)}| \,, \tag{5.10}$$

where N is the batch size, and $Y_{(n,w,h)}$ and $\hat{Y}_{(n,w,h)}$ are the predicted and observed images, respectively.

In addition, to increase local variability and balance the blurring effects from Eq. (5.10), we use a perceptual loss (*Ledig et al.*, 2017; *Johnson et al.*, 2016) based on a pre-trained VGG-19 network (*Simonyan and Zisserman*, 2015) on ImageNet (*Deng et al.*, 2009). This additional loss constrains the generated images to have a similar structure and spatial variability to the observed images by comparing multi-level features extracted by a VGG classifier network from both the predicted and observed images:

$$\mathcal{L}_{VGG} = 8\mathcal{L}^1_{VGG} + \sum_{j=2}^{J} \mathcal{L}^j_{VGG} \,, \tag{5.11}$$

$$\mathcal{L}^j_{VGG} = \frac{1}{NC^jW^jH^j} \sum_{n=1}^{N} \sum_{c=1}^{C^j} \sum_{w=1}^{W^j} \sum_{h=1}^{H^j} |\phi^j(\mathbf{Y}_{(n,j,c,w,h,)}) - \phi^j(\hat{\mathbf{Y}}_{(n,j,c,w,h)})| \,, \tag{5.12}$$

where J is the number of levels from which the VGG features are extracted, W^j and H^j are the spatial extensions of the respective level within the VGG classifier, C^j is the number of channel dimension of the respective level, and $\phi^j(\mathbf{Y}_{(n,j,c,w,h,)})$ and $\phi^j(\hat{\mathbf{Y}}_{(n,j,c,w,h,)})$ are the extracted features at level j from the predicted and observed images, respectively. In contrast to classification problems where high level features play a more important role, we multiply the low level features by a weighting factor of 8 to preserve the local features and give them more importance since these are more relevant to our regression task. The VGG network was originally trained with RGB images and giving NDVI and BT as input is not directly possible. To solve this issue, we replicate NDVI and BT along the channel dimension and feed each of them separately to the VGG network. The impact of using this perceptual loss is evaluated in Section 5.3.5.1. The entire loss function to be minimized is thus given as follows:

$$\mathcal{L} = \mathcal{L}^{NDVI}_{MAE} + 0.1\mathcal{L}^{NDVI}_{VGG} + \mathcal{L}^{BT}_{MAE} + 0.1\mathcal{L}^{BT}_{VGG} \,, \tag{5.13}$$

where $\mathcal{L}^{NDVI}_{MAE}$ and $\mathcal{L}^{NDVI}_{VGG}$ are the MAE and VGG losses on NDVI and $\mathcal{L}^{BT}_{MAE}$ and $\mathcal{L}^{BT}_{VGG}$ are the MAE and VGG losses on BT, respectively. The weighting factor 0.1 is set to balance the losses. More technical details regarding the training are provided in the next Section 5.3.1.2.

5.3 Experimental results and analysis

This section is organized as follows. Section 5.3.1 outlines the experimental setup including the baselines approaches. Results on predicting NDVI and BT are given in Section 5.3.2. Section 5.3.3 discusses an application of the model for agricultural drought assessment. Analysis about the TSMP model components and variables is given in Section 5.3.4. Finally, ablation studies are described Section 5.3.5.

5.3.1 Experimental Setup

5.3.1.1 Baseline Approaches

We study the performance of recently developed vision transformers on our task. We achieve this by sharing the overall model architecture and implementing the main building block inside the encoder and decoder according to different algorithms. The implemented models are as follows:

U-Net (*Ronneberger et al.*, 2015). This serves as a baseline of typical U-Net models. We implemented this model based on a 2D CNN with residual convolutional blocks. The U-Net model does not use an attention mechanism.

Swin Transformer V1 (*Liu et al.*, 2021). This model performs self-attention in shifted windows to reduce the computational complexity compared to the original ViT. Transformers based on this model have been commonly applied for variety of tasks in remote sensing and computer vision (*Wang et al.*, 2022c; *Gao et al.*, 2021; *Wang et al.*, 2022d; *Aleissaee et al.*, 2023).

Swin Transformer V2 (*Liu et al.*, 2022a). This is an improved model of Swin V1. The attention mechanism is replaced with a scaled cosine attention to measure pixel feature similarities. Swin V2 utilizes post normalization layers inside the main block thus making the optimization of large models more stable. In addition, it proposes to replace the positional encoding inside the windows with a log-spaced continuous one to ease downstream tasks with pre-trained models.

Wave-MLP (*Tang et al.*, 2022a). This is an MLP-Mixer-based transformer model. The basic block is built on a stack of MLPs. Wave-MLP represents each pixel as a wave function with amplitude features representing pixel contents and phase to measure the relations with other pixels.

Climatology. Apart of these models, we report the results for two NDVI/BT climatology baselines. The climatology is based on multi-year mean values computed from remote sensing observations pixel-wise and on a weekly basis. The first is a **climatology-I** computed from the years 1981-1988 which represents a prescribed satellite phenology before the beginning of the simulation. The second is a **climatology-II** computed from the training years 1989-2016 in an overlap with the simulation period. The later climatology represents a function that models the annual cycles and it can be used to check if the models generalize beyond the mean annual cycles of the predicted NDVI/BT.

5.3.1.2 Implementation Details

We implemented all aforementioned DL models in our framework and trained them with 3 different random seeds, this ensures a fair comparison and better estimation. All models have almost the same capacity with $\sim$ 12 million parameters. The encoders for the transformer models were pre-trained on ImageNet-1K (*Deng et al.*, 2009) while the weights in the decoders and regression heads were initialized randomly from a standard normal distribution. To increase generalization and robustness of the models, we use 4 augmentation techniques. This includes flipping and rotating of the input with a probability of 0.5 and randomly perturbing the input variables by adding noise from a normal distribution with zero mean and a standard deviation of 0.02 with a probability of 0.5. In addition,

to generate the input corresponding to week i during training, we randomly average two days corresponding to the week i as an additional augmentation technique. All models were trained with the $\mathcal{L}$ loss Eq. (5.13) using the Pytorch framework (*Paszke et al.*, 2019) with a learning rate 0.0003 and a scheduler to decay the learning rate by a factor of 0.9 every 16 epochs. AdamW optimizer (*Loshchilov and Hutter*, 2019) was used for the gradient descent with ($\beta_1 = 0.9,\ \beta_2 = 0.999$) and a weight decay 0.05. We use dropout probability of 0.2 and a stochastic depth rate of 0.3. We train with a batch size of $N = 2$ for 100 epochs. For Swin Transformers, we set the window size to 8 and use the following number of heads $\{3, 6, 12\}$ for the encoder and the same order for the decoder. The down sampling in the encoder followed the original implementation in Swin Transformer. Wave-MLP was trained with the dimensionality $\{C_{(en,1)} = 64, C_{(en,2)} = 128, C_{(en,3)} = 320\}$ and $r_{mlp} = 4$ for both the encoder and the decoder. Wave-MLP and Swin V2 use a dropout probability of 0.1 and a stochastic depth rate of 0.2. In addition, we follow the official implementation of Wave-MLP and use GroupNorm (*Wu and He*, 2018) with a group of 1 instead of LayerNorm.

For the experiments about channel attention in Section 5.3.4, we trained two separated models for NDVI and BT independently with ($r_{att} = 3$, $r_{att} = 5$) and with the dimensionality $\{C_{(en,1)} = 64, C_{(en,2)} = 128, C_{(en,3)} = 256\}$, and averaged the results.

Finally, all models were trained on individual NVIDIA RTX A6000 GPUs with 48 GB. Additional details and code are available at `https://github.com/HakamShams/Focal-TSMP`.

5.3.1.3 Dataset

We use the TSMP simulation described in Section 2.2.2.1 as input. As a target for training and evaluation, we use the observed remote sensing data described in Section 2.2.3. The preprocessing framework of the TSMP simulation is describe in Section 2.2.2.2 and the preprocessing of the remote sensing data is described in Section 2.2.3.2. We reserved the years 1989-2009 and 2013-2016 for training, 2010-2011 and 2017 for validation, and 2012, 2018-2019 for testing (Section 2.2.2.2).

5.3.1.4 Evaluation Metrics

To measure the model performance, we use the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Bias, coefficient of determination (R2), Pearson Correlation Coefficient (R_p), and Spearman Correlation Coefficient (R_s). We compute the metrics for each sample and then average the values to obtain the final metrics.

5.3.2 NDVI and BT Prediction

The quantitative results of the models are shown in Tables 5.1 and 5.2. Pixels without a vegetation cover (e.g., pixels over desert) were excluded from the results. Including these pixels will overestimate the model performance since they have small variations throughout the years. For the masking, we use NOAA quality assurance (QA) metadata. As can be seen from Tables 5.1 and 5.2, all DL models outperform the first climatology-I baseline with a huge margin. This is because the climatology was calculated before the simulation run. This climatology can not capture the dynamic after 3 decades. The second climatology-II baseline is stronger. It uses information from multiple years within the simulation run. All DL models still achieve better results indicating that the models have learned the seasonal dynamic beyond climatology. In addition, these climatologies baselines can not

be used to derive drought indices (Section 5.3.3) since the inter-annual variability in NDVI and BT is neglected as average cycles are used. Furthermore, comparing the correlation and results of BT with NDVI, we can observe that all models achieve higher correlation metrics (R2, R_p, and R_s) on BT than NDVI. This can be explained by the fact that NDVI is a composition of two bands while BT is only derived from the infrared band, thus it is harder for the models to estimate NDVI than BT. In general, all DL models provide close results and are considered suitable for the task. Focal Modulation clearly outperformed other DL models on the validation set for both NDVI and BT predictions. For the test set on NDVI, it comes slightly after the Wave-MLP model. However, Focal Modulation can generalize better for BT thus providing a balanced prediction between NDVI and BT and consequently it is capable to generate an overall better prediction.

In Table 5.3, we report the estimated inference time for the DL models. For the Focal Modulation model, the estimated inference time to generate one sample for NDVI and BT containing $397 \times 409 \times 2$ grid points is 0.24 ± 0.01 seconds on one NVIDIA GeForce RTX 3090 GPU and 12 ± 0.1 seconds on one AMD Ryzen 9 3900X 12-Core CPU. U-Net with 2D CNN dose not include operations for the attention mechanism thus it is the fastest but the performance is lower.

Qualitative results for the model prediction with Focal Modulation are shown in Figs. 5.5 and 5.6. We take weeks from different seasons through the years and remove pixels over desert for the calculations of bias distribution and regression line. Positive bias values mean that the model overestimates NDVI (BT) while negative ones indicate that the model underestimates NDVI (BT).

Validation - Years (2010, 2011, 2017) - 156 weeks					
NDVI					
Algorithm	MAE(↓)	RMSE(↓)	R2(↑)	R_p(↑)	R_s(↑)
Climatology-I	0.0550	0.0680	0.5763	0.8939	0.8669
Climatology-II	0.0326	0.0416	0.8372	0.9353	0.9113
U-Net	0.0277±0.0001	0.0365±0.0002	0.8743±0.0008	0.9406±0.0005	0.9172±0.0005
Wave-MLP	<u>0.0272</u>±0.0003	**0.0358**±0.0003	<u>0.8784</u>±0.0018	<u>0.9422</u>±0.0018	<u>0.9183</u>±0.0021
Swin Transformer V1	0.0273±0.0003	0.0362±0.0003	0.8759±0.0022	0.9411±0.0013	0.9161±0.0023
Swin Transformer V2	0.0277±0.0003	0.0369±0.0003	0.8703±0.0021	0.9415±0.0010	0.9167±0.0008
Focal Modulation	**0.0269**±0.0001	**0.0358**±0.0002	**0.8790**±0.0017	**0.9432**±0.0001	**0.9194**±0.0009
BT (K)					
Algorithm	MAE(↓)	RMSE(↓)	R2(↑)	R_p(↑)	R_s(↑)
Climatology-I	2.9130	3.7302	0.8454	0.9466	0.9408
Climatology-II	2.3017	3.0020	0.8963	0.9601	0.9539
U-Net	1.9377±0.0093	2.6067±0.0057	0.9243±0.0014	0.9667±0.0004	<u>0.9603</u>±0.0007
Wave-MLP	<u>1.9200</u>±0.0491	<u>2.5834</u>±0.0486	<u>0.9248</u>±0.0035	<u>0.9668</u>±0.0006	<u>0.9603</u>±0.0007
Swin Transformer V1	1.9642±0.0246	2.6341±0.0303	0.9221±0.0012	0.9661±0.0005	0.9590±0.0006
Swin Transformer V2	1.9741±0.0191	2.6420±0.0258	0.9225±0.0013	0.9659±0.0011	0.9590±0.0014
Focal Modulation	**1.9010**±0.0071	**2.5364**±0.0073	**0.9280**±0.0012	**0.9679**±0.0001	**0.9614**±0.0007

Table 5.1: Comparing the performance of different DL models. The metrics are shown for the validation set. The best and second best results on each metric are highlighted in bold and underlined text, respectively. (±) denotes the standard deviation for 3 different runs.

Test - Years (2012, 2018, 2019) - 139 weeks					
NDVI					
Algorithm	MAE(↓)	RMSE(↓)	R2(↑)	R_p(↑)	R_s(↑)
Climatology-I	0.0567	0.0697	0.5529	0.8933	0.8704
Climatology-II	0.0314	0.0400	0.8507	0.9433	0.9254
U-Net	0.0274±0.0004	0.0359±0.0005	0.8772±0.0006	0.9435±0.0006	0.9237±0.0009
Wave-MLP	**0.0261**±0.0006	**0.0343**±0.0008	**0.8861**±0.0043	**0.9467**±0.0024	0.9252±0.0011
Swin Transformer V1	0.0269±0.0003	0.0355±0.0004	0.8795±0.0029	0.9442±0.0010	0.9239±0.0014
Swin Transformer V2	0.0270±0.0005	0.0359±0.0005	0.8766±0.0038	0.9447±0.0012	0.9251±0.0020
Focal Modulation	0.0266±0.0003	0.0350±0.0004	0.8808±0.0014	0.9454±0.0009	**0.9253**±0.0016
BT (K)					
Algorithm	MAE(↓)	RMSE(↓)	R2(↑)	R_p(↑)	R_s(↑)
Climatology-I	2.8806	3.6864	0.8447	0.9485	0.9470
Climatology-II	2.2024	2.8880	0.9036	0.9623	0.9606
U-Net	1.9920±0.0148	2.6652±0.0262	0.9164±0.0021	0.9644±0.0009	0.9616±0.0005
Wave-MLP	1.9376±0.0184	2.6221±0.0177	0.9172±0.0005	0.9647±0.0005	0.9619±0.0008
Swin Transformer V1	1.9563±0.0329	2.6381±0.0397	0.9169±0.0038	0.9649±0.0009	0.9627±0.0008
Swin Transformer V2	1.9516±0.0639	2.6277±0.0874	0.9183±0.0060	0.9641±0.0025	0.9619±0.0020
Focal Modulation	**1.9179**±0.0458	**2.5745**±0.0470	**0.9204**±0.0030	**0.9664**±0.0007	**0.9636**±0.0006

Table 5.2: Comparing the performance of different DL models. The metrics are shown for the test set. The best and second best results on each metric are highlighted in bold and underlined text, respectively. (±) denotes the standard deviation for 3 different runs.

Algorithm	GPU[1]	CPU[2]
U-Net	0.09±0.02	5±0.2
Wave-MLP	0.28±0.00	10±0.3
Swin Transformer V1	0.18±0.00	11±0.2
Swin Transformer V2	0.19±0.00	11±0.2
Focal Modulation	0.24±0.01	12±0.1

[1]NVIDIA GeForce RTX 3090 GPU
[2]AMD Ryzen 9 3900X 12-Core CPU

Table 5.3: Inference time in seconds for different DL models.

As shown in Figs. 5.5 and 5.6, the biases vary across the weeks and locations. For week 7 in 2012, the biases for both NDVI and BT are relatively high. Week 26 in 2019 exhibits similar high biases in both NDVI and BT over high latitudes regions. The respective distribution of biases is also shown in Figs. 5.5 and 5.6. Overall, the results show that the dynamics over the years are well captured. The biases for both NDVI and BT are closely centered around zero with a shift for the center of bias distribution from zeros. This shift is however in the same direction for both NDVI and BT. We can also observe that the model fits the regression lines better for weeks 14, 26, and 39 than for week 7 in winter 2012. The comparison between the distributions of predicted and observed NDVI and BT confirms also the observation that the model captured the dynamic throughout the years.

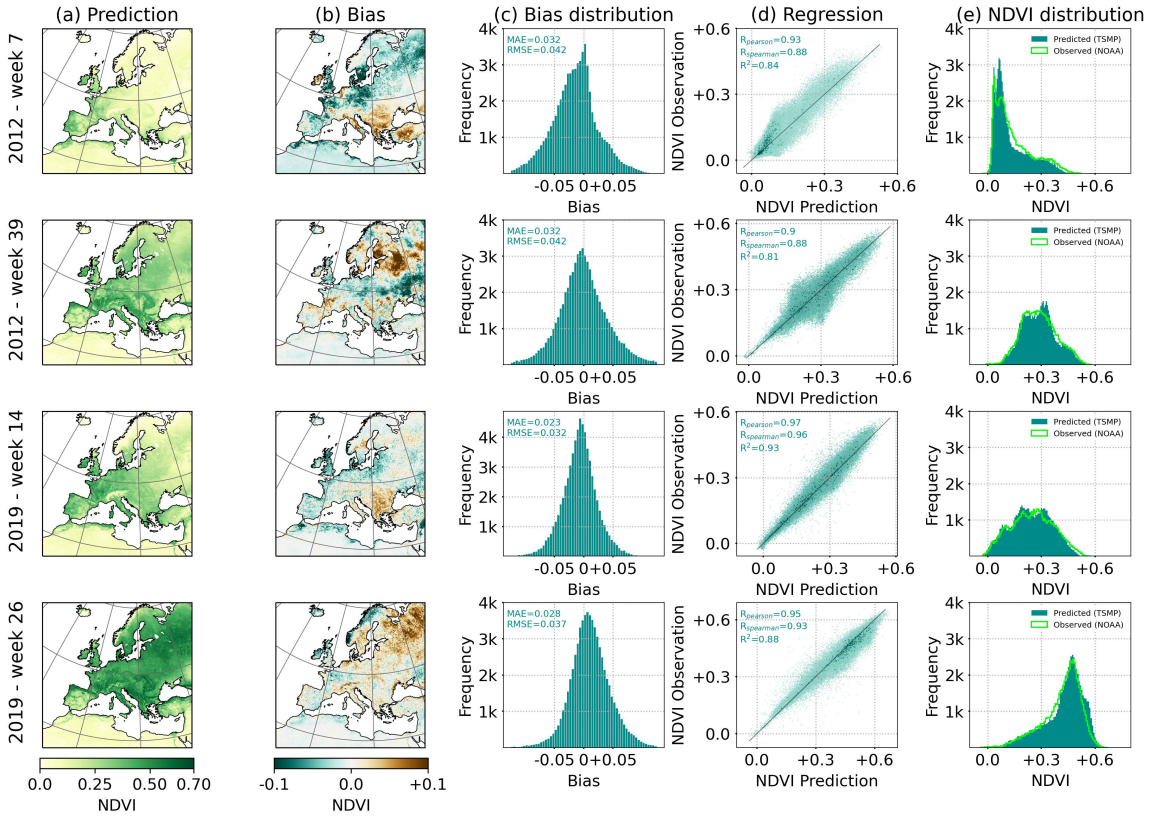


Figure 5.5: Examples predictions for weekly NDVI from the test set. (a) Predicted NDVI. (b) Bias computed as prediction minus observed. (c) Distribution of biases. (d) Regression results as predicted versus observed. (e) Distribution of NDVI values for NOAA observation and model prediction. The metrics are computed over all pixels with vegetation cover.

While this provides examples of the performance for individual samples, in Fig. 5.7 we provide an additional experiment where we analyze biases of model predictions within different seasons of the year and over PRUDENCE regions (see Section 2.4, Fig. 2.10 for the definition of PRUDENCE regions). This allows us to assess the model weaknesses and strengths with different seasonality and spatial variability. The mean biases were computed pixel-wise from both the validation and test years time-series where we computed the biases for each pixels from the weeks that belong to a specific season and averaged the results to obtain the last metric. In addition, we computed Pearson Correlation R_p pixel-wise in a similar way. As seen in Fig. 5.7, there are clusters of positive/negative biases that vary with seasons over specific regions. For instance, for NDVI prediction, the eastern part of British Isles exhibits positive biases for all seasons while Iceland and north Africa show constant negative biases. For BT, Southeastern Europe has persistent positive biases with larger errors during winter. Pixels over desert, e.g., north Africa, show less variability in NDVI where only little seasonality is shown as in Fig. 5.5. Thus, such regions are easier to predict with relatively small biases. However any fluctuation in NDVI prediction over these pixels will lead to lower correlation compared to other regions since the time series primarily represent small variations around the mean NDVI value. In comparison to other seasons, the winter season has relatively poor predic-

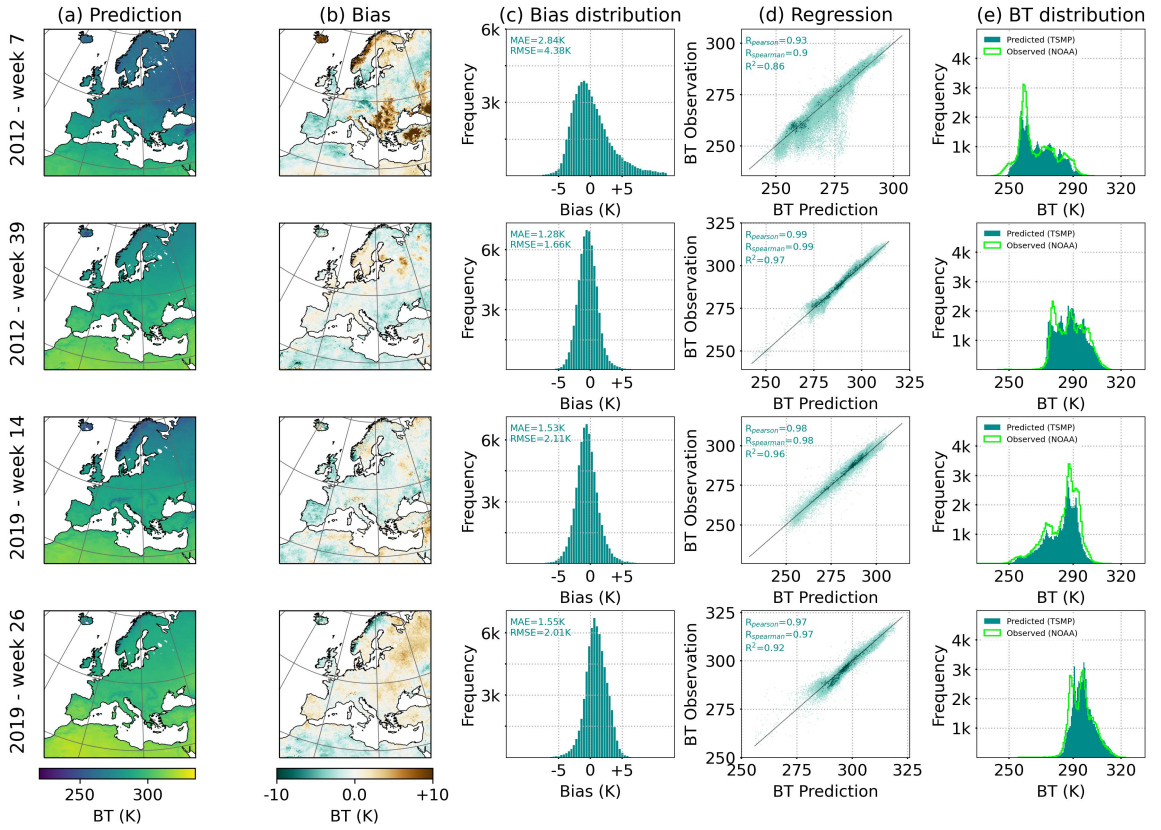


Figure 5.6: Examples predictions for weekly BT from the test set. (a) Predicted BT. (b) Bias computed as prediction minus observed. (c) Distribution of biases. (d) Regression results as predicted versus observed. (e) Distribution of BT values for NOAA observation and model prediction. The metrics are computed over all pixels with vegetation cover.

tions especially in the high latitudes regions. One possible explanation for these errors is the lack of accurate training data in Scandinavian regions during winter. For instance, previous studies on ParFlow-CLM models showed that hydrological modeling performs worse in northeastern Europe due to errors in snow dynamics and regional forces (*Naz et al.*, 2023; *Furusho-Percot et al.*, 2019). It was also shown by *Yang et al.* (2020) and *Eisfelder et al.* (2023) that high latitude regions are less reliable to derive vegetation products due to snow cover and its effects on the albedo and larger sensor zenith angles. Another source of model errors is that NOAA vegetation products depend on temporal compositing to handle high frequency and atmosphere transmittance (*Yang et al.*, 2020). The absence of a generalized physical-based model to enhance accuracy over various surfaces and for all conditions generates difficulties for satellite products (*Kogan*, 1995a). *Nagol et al.* (2009) assessed the uncertainty of NDVI in this regards. These issues add some uncertainties to the model training and evaluation. Using more recent atmospheric correction methods such as in (*Moravec et al.*, 2021) could also enhance the results. Furthermore, as mentioned in Section 2.2.2.1, the TSMP simulation was performed in a free mode and had no modelling of anthropogenic-related influences. Given that agricultural systems and human activities which are interlinked with drought events could change and follow adaptation strategies (*Van Loon et al.*, 2016a), this certainly contributes to the

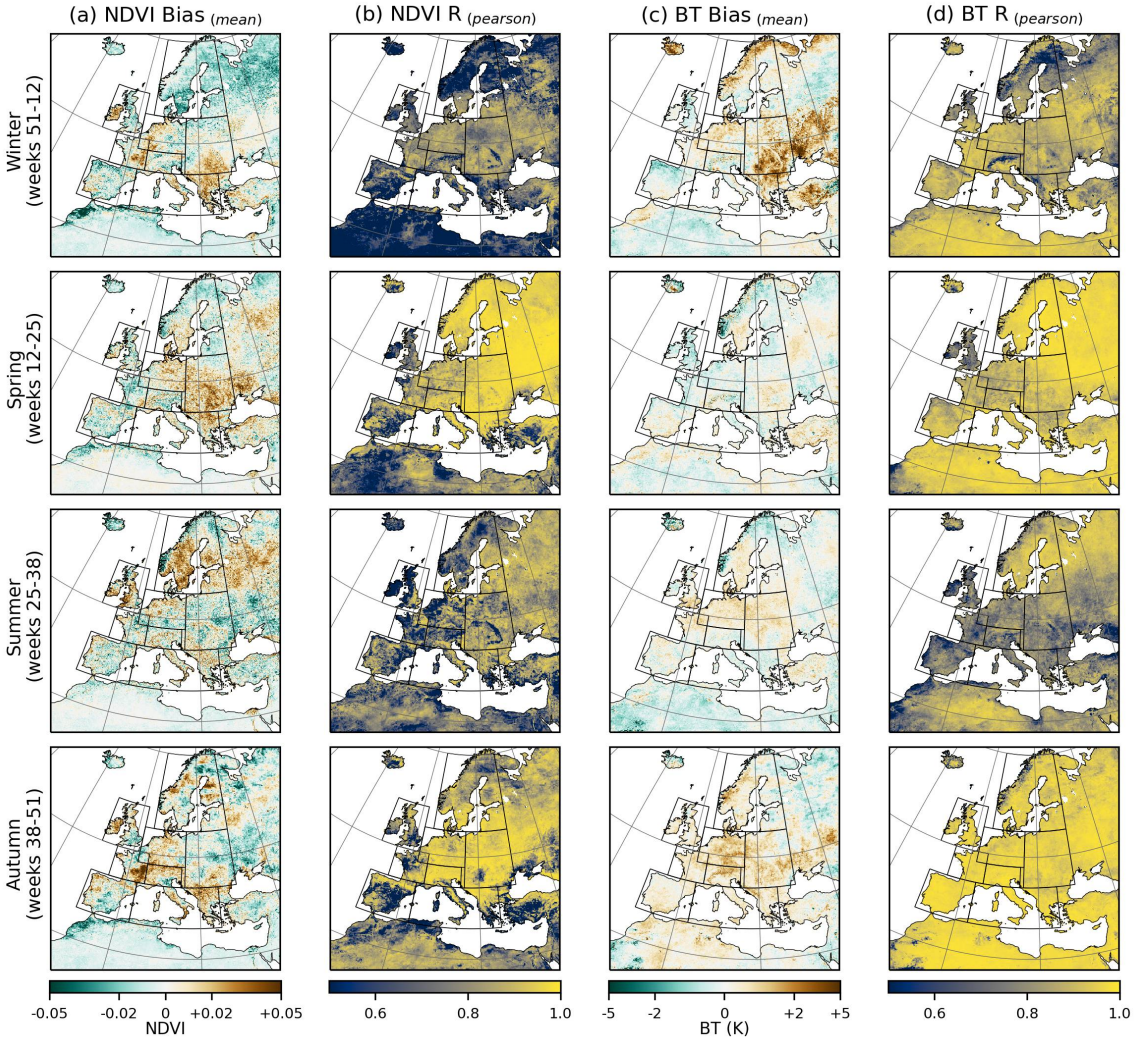


Figure 5.7: An analysis of uncertainty and model generalization for different times of the year. The analysis was performed on the validation and test sets as one set. (a) NDVI mean bias. (b) NDVI mean Pearson Correlation. (c) BT mean bias. (d) BT mean Pearson Correlation.

error budget of the model. Developing realistic land use and water management scenarios within a probabilistic TSMP could reduce these errors. In addition, the uncertainty in TSMP is highly linked to potential errors in the driving forces and spin-up initialization. While these errors are common limitations of simulations and remote sensing data, it should be noted that the prediction of a DL model has its own uncertainty. Therefore, more efforts are needed to recognize the sources of uncertainty in model prediction (Section 5.3.3).

In Fig. 5.8, we visualize the computations over each PRUDENCE region separately. For Figs. 5.8 (a) and 5.8 (b), we fit a normal distribution over the normalized histogram of biases for each season and over all PRUDENCE regions. For instance, positive shifts of the estimated means are shown in NDVI for both FR and AL regions during autumn. The same pattern is shown for SC and BI during summer. As can also be seen in Fig. 5.8 (b), a positive shift for BT is shown for all regions during

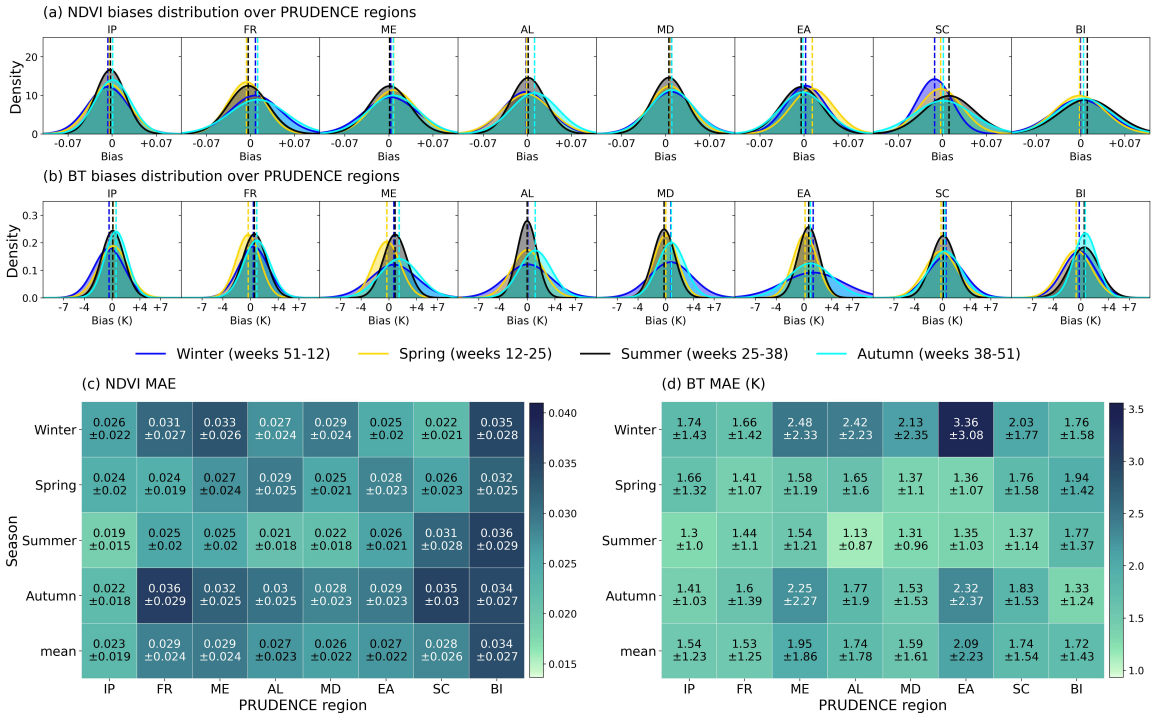
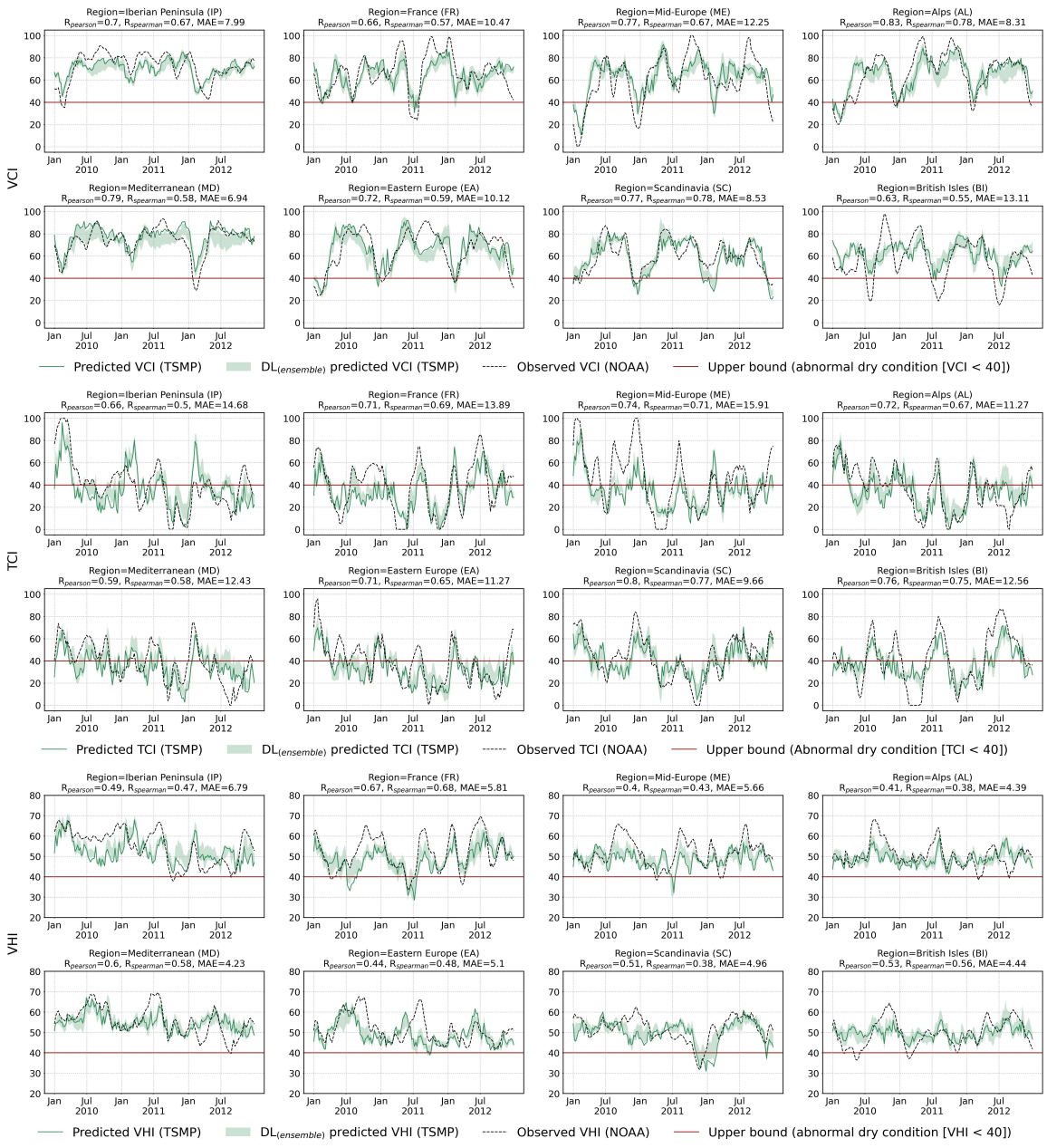


Figure 5.8: An analysis of uncertainty and model generalization for different time of the year over each PRUDENCE region. The analysis was performed on the validation and test sets as one set. (a) NDVI bias distribution. (b) BT bias distribution. Shown are the probability density functions. (c) NDVI MAE. (d) BT MAE.

autumn. Furthermore, the shape of the distribution gives an overview of the prediction homogeneity within the region, e.g., the prediction is highly uncertain over EA during winter and consequently has a relatively high standard deviation. The mean values in Figs. 5.8 (c) and 5.8 (d) represent the expected MAE for all seasons combined. Fig. 5.8 (c) indicates that in general the model predictions for NDVI are less certain during autumn in comparison to other periods and over BI within the PRUDENCE regions. For BT, it can be seen from Fig. 5.8 (d) that the prediction is less certain during winter and over ME and EA regions.



. NDVI and BT anomalies are provided in Appendix Fig. 5.11.

Figure 5.9: Comparison of spatially averaged weekly agricultural drought indices between the model prediction and NOAA observation over each PRUDENCE region. Drought indices were computed from the long-term climatology (1989-2016) pixel-wise and on a weekly basis. All results are obtained with the Focal Modulation Network. The ensemble model is the result of all DL models described in Section 5.3.1.1

Figure 5.10: Comparison of spatially averaged weekly agricultural drought indices between the model prediction and NOAA observations over each PRUDENCE region. Drought indices were computed based on the long-term climatology (1989-2016) pixel-wise and on a weekly basis. All results are obtained with the Focal Modulation Network. The ensemble model is the result of all DL models described in Section 5.3.1.1.

5.3.3 Agricultural Drought Assessment

In this section, we assess the model capability to predict different agricultural droughts indices on a high temporal resolution (weekly basis). More specifically, we use the predicted NDVI and BT along with their multi-year climatology to derive NDVI anomaly, BT anomaly, VCI, TCI, and VHI drought indices. NDVI and BT anomalies were computed by subtracting the mean value of the respective pixel and week from the predictions (observations). VCI, TCI, and VHI were computed from Eq. (2.21)-(2.23). Figs. 5.9 and 5.10 compare the predicted agricultural drought indices VCI, TCI and VHI by the focal modulation model with the observed ones from NOAA remote sensing data for the years 2010-2012 (Fig. 5.9) and 2017-2019 (Fig. 5.10). We spatially average the values inside each PRUDENCE region and plot their respective time-series on a weekly basis. Generally, values below 40 are identified as abnormally dry conditions (*Kogan et al.*, 2015; *Yang et al.*, 2020).

Overall, the prediction resembles the seasonal wetness and dryness at the regional scale. The agreements between predictions and observations vary across regions and time with satisfactory R_p

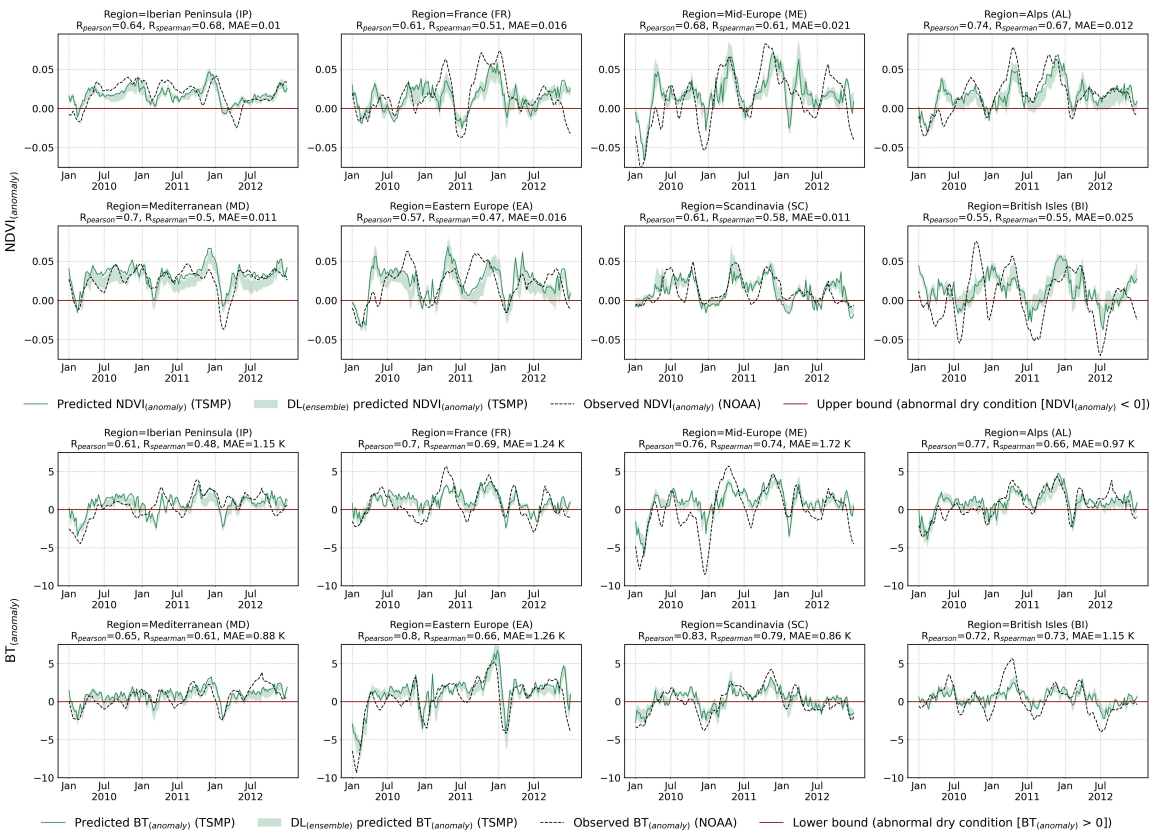


Figure 5.11: Supplementary results to Fig. 5.9. Comparison of spatially averaged weekly NDVI anomalies between the model prediction and NOAA observation over each PRUDENCE region. Anomaly was computed by subtracting the mean values from predictions (observations). The mean values were computed from the long-term climatology (1989-2016) pixel-wise and on a weekly basis. All results are obtained with a DL model based on Focal Modulation Network. The ensemble model is the result of all DL models described in Section 5.3.1.1.

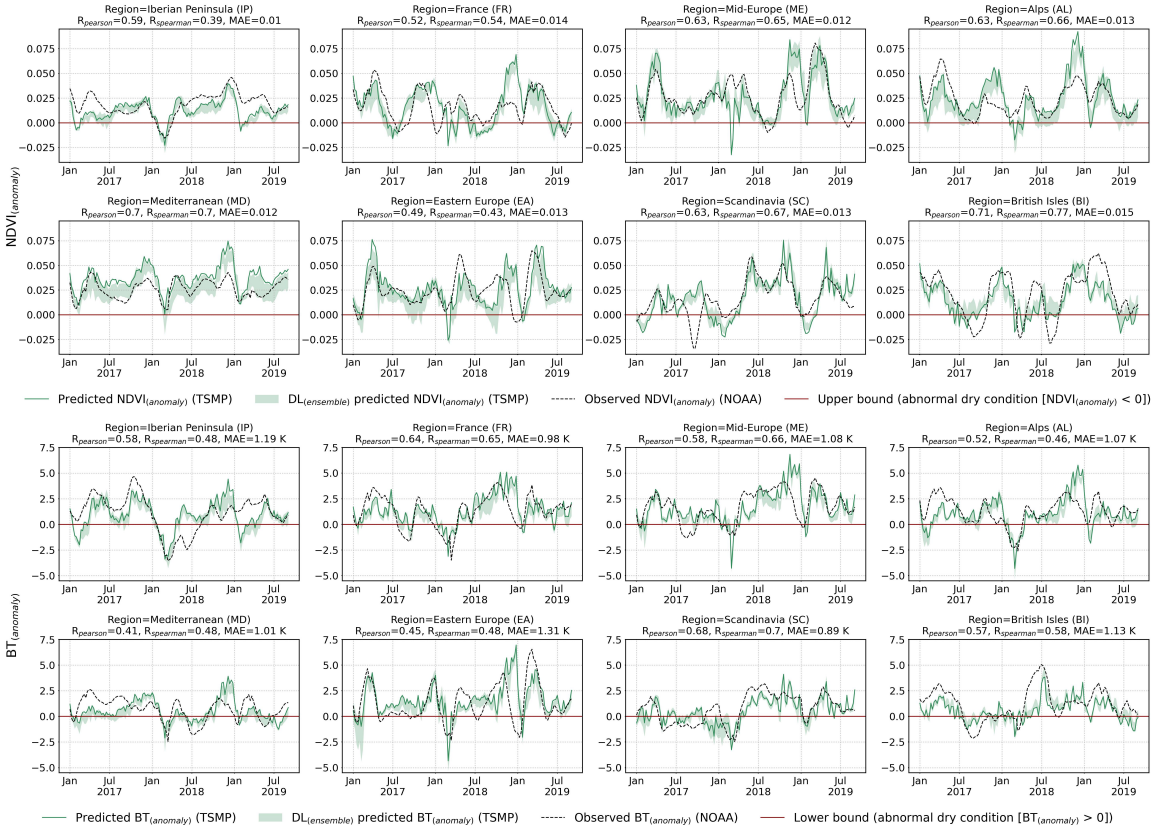


Figure 5.12: Supplementary results to Fig. 5.10. Comparison of spatially averaged weekly BT anomalies between the model prediction and NOAA observation over each PRUDENCE region. Anomaly was computed by subtracting the mean values from predictions (observations). The mean values were computed from the long-term climatology (1989-2016) pixel-wise and on a weekly basis. All results are obtained with a DL model based on Focal Modulation Network. The ensemble model is the result of all DL models described in Section 5.3.1.1.

values ranging from 0.50 to 0.77, 0.38 to 0.70, and 0.50 to 0.75 for VCI, TCI, and VHI, respectively. MAE values fluctuate in the range 9.99-6.81, 13.88-10.24, and 5.80-2.69 for VCI, TCI, and VHI, respectively. While there is a satisfactory agreement with observations, there are some obvious discrepancies, e.g., in TCI over the Iberian Peninsula (IP) during summer 2018. More interestingly, we show the bounded results of an ensemble of DL models. This ensemble is the results of all DL models. As can be seen, all DL models which are based on different algorithms yield close predictions with small standard deviations. This supports that errors in model prediction are probably to be more attributed to biases in the TSMP model and remote sensing reference data. In this respect, *Yang et al.* (2021b) showed that vegetation products over regions with extreme little seasonality, e.g., desert and high mountains have higher errors. This can be seen from Eq. (2.21)-(2.23), where small differences between maximum and minimum values could lead to higher deviation in the vegetation indices.

Finally, as observed from the plots, the thermal surface condition represented by TCI contributes more to the agricultural drought events over Europe than the deficiency in vegetation moisture condition approximated as VCI. This is in agreement with (*Zeng et al.*, 2023), who showed that drought

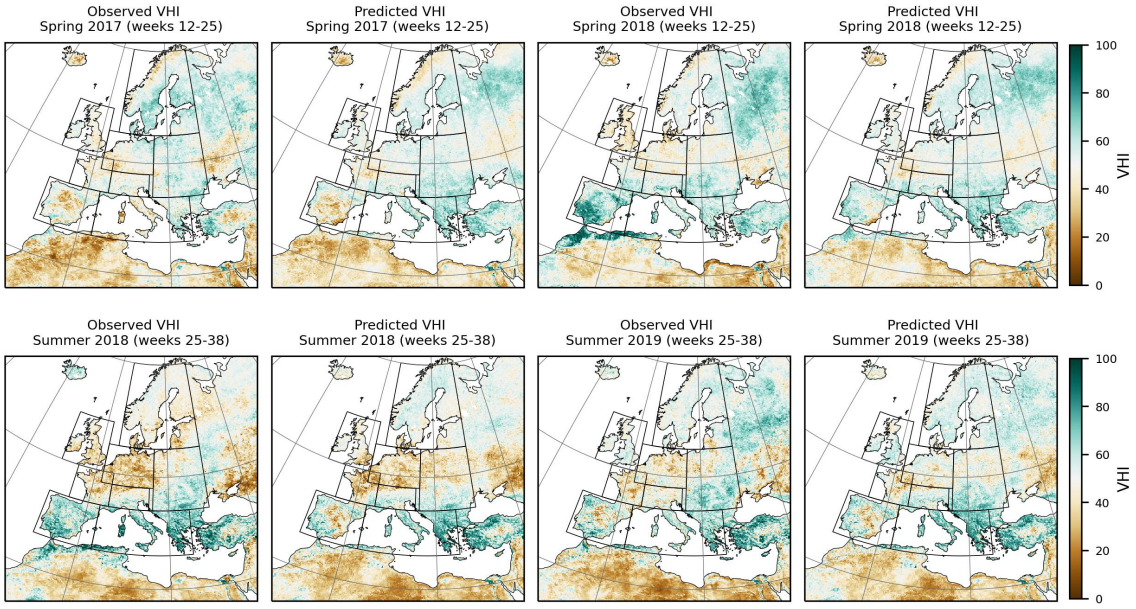


Figure 5.13: Comparison between the seasonal predicted Vegetation Health Index (VHI) and NOAA observations over Pan-Europe domain.

affecting vegetation is more likely to be associated with high abnormal temperatures in Europe. This is critical for studies that rely on NDVI as the solely vegetation product to identify drought events over Europe (Section 3.2.2). In Figs. 5.11 and 5.12, we show the time-series for NDVI and BT anomalies. We also show vegetation health maps for different seasons from the validation and test years. These predicted maps are depicted in Fig. 5.13. As shown, the model predicts an increase of agricultural droughts in the summer of 2018 in the Mid Europe and France regions. *Xoplaki et al.* (2023) associated this extremely dry summer with compound extreme events.

Furthermore, in Fig. 5.14, we provide an analysis about the frequency of extreme droughts for the two periods 2010-2012 and 2017-2019. Frequency represents the percent of weeks with Severe-to-Exceptional drought events where VHI < 26 (*Kogan et al.*, 2020).

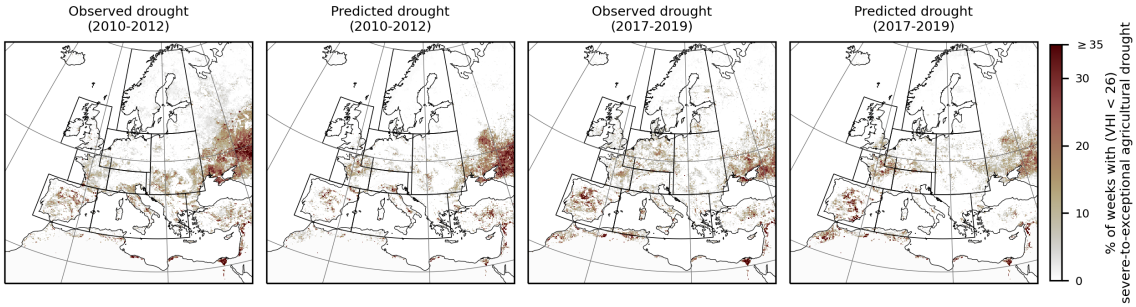


Figure 5.14: Comparison between the predicted drought frequency and NOAA observations over Pan-Europe domain. Frequency represents the percent of weeks with Severe-to-Exceptional drought events (VHI < 26).

While Figs. 5.9 and 5.10 provide overviews of the averaged values over the regions, the analysis in Fig. 5.14 provides a spatial comparison between the model prediction and observations. The major hotspots for the highest extremes are found outside the Prudence regions (North of the Black Sea, Northwest Africa, Egypt and Northwest of the Middle East). In comparison to the Prudence regions, the Iberian Peninsula and France exhibit more extreme droughts. The model predicts more extreme droughts in those regions and agrees with observations. For the period 2010-2012, the model predicts less extreme droughts in the Mediterranean and Eastern Europe. While for the 2017-2019 period, the model underestimates the frequency of extremes in the Mid Europe region.

Moreover, Fig. 5.15 evaluates the model capability to capture seasonal dynamic in drought indices. As seen in Fig.5.15 (a), the mean R_p values are greater than 0.5 and around 0.6 for all seasons. MAE values show the highest error in VCI for the winter season. One notable observation is that the error bars have relatively large values indicating a variation in prediction accuracy across the years within the same seasons. This can be attributed to the seasonality shift in the long-term trends. *Klimavičius et al.* (2023) showed that meteorological forces like air temperature have strong impact on growing seasons and phenological trends of NDVI (VCI). The cumulative distribution functions (CDF) in Fig. 5.15 (b) expresses the main difference in CDF for VCI during winter. While the model prediction overestimates TCI over the seasons.

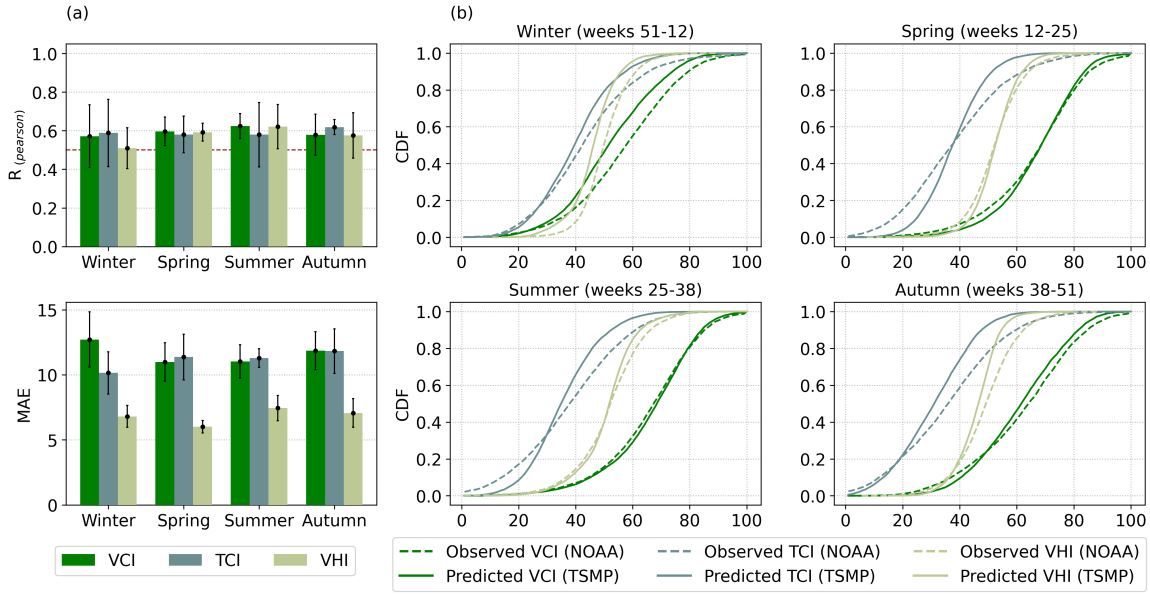


Figure 5.15: An evaluation of seasonally predicted agricultural drought indices with ground truth NOAA observations at the resolution 0.88°. (a) Bottom is mean absolute errors (MAE) and top is Pearson Correlations (R_p) for different seasons. (b) Comparison of the cumulative distribution functions between prediction and observations.

5.3.4 Variable Importance

To analyze the impact of each TSMP model components on the model prediction, we present in Table 5.4 the prediction results obtained with COSMO, CLM, and ParFlow. For this experiment, we train 3

models based on focal modulation with the dimensionality $\{C_{(en,1)} = 64, C_{(en,2)} = 128, C_{(en,3)} = 256\}$. As seen in Table 5.4, compared to CLM and ParFlow, COSMO achieves the best results for the validation set while CLM outperforms both for the test set. COSMO has important variables related to water contents and clouds along with other variables related to the atmospheric effects on the reflected signal on the ground. CLM has complementary variables related to heat fluxes and evapotranspiration. ParFlow can approximate the hydrology and serve as a proxy for the soil conditions. The results show that all model components are useful and the best result is obtained when all these models are used.

While Table 5.4 provides an overview on the importance of model components, apriori choice of proper input variables from each of these model components to predict NDVI and BT requires

Validation - Years (2010, 2011, 2017) - 156 weeks					
NDVI					
Model	MAE(↓)	RMSE(↓)	R2(↑)	R_p(↑)	R_s(↑)
COSMO	0.0281	0.0372	0.8696	0.9403	0.9160
CLM	0.0289	0.0382	0.8586	0.9369	0.9115
ParFlow	0.0303	0.0396	0.8500	0.9314	0.9042
COSMO + CLM + ParFlow	**0.0270**	**0.0359**	**0.8781**	**0.9433**	**0.9184**
BT (K)					
Model	MAE(↓)	RMSE(↓)	R2(↑)	R_p(↑)	R_s(↑)
COSMO	1.9975	2.6389	0.9227	0.9667	0.9615
CLM	2.0187	2.7080	0.9160	0.9653	0.9600
ParFlow	2.2029	2.9254	0.9052	0.9617	0.9545
COSMO + CLM + ParFlow	**1.8981**	**2.5433**	**0.9266**	**0.9679**	**0.9613**
Test - Years (2012, 2018, 2019) - 139 weeks					
NDVI					
Model	MAE(↓)	RMSE(↓)	R^2(↑)	R_p(↑)	R_s(↑)
COSMO	0.0285	0.0372	0.8619	0.9437	0.9238
CLM	0.0269	0.0355	0.8782	0.9443	0.9238
ParFlow	0.0291	0.0379	0.8648	0.9396	0.9175
COSMO + CLM + ParFlow	**0.0268**	**0.0353**	**0.8795**	**0.9452**	**0.9243**
BT (K)					
Model	MAE(↓)	RMSE(↓)	R^2(↑)	R_p(↑)	R_s(↑)
COSMO	2.0847	2.7549	0.9060	0.9633	0.9612
CLM	1.9362	2.6303	0.9185	0.9650	0.9637
ParFlow	2.2663	2.9481	0.8962	0.9635	0.9604
COSMO + CLM + ParFlow	**1.8730**	**2.5277**	**0.9227**	**0.9672**	**0.9642**

Table 5.4: Impact of TSMP model components on the model performance. The metrics are shown for the validation and test sets. All models were trained with Focal Modulation Network. The best results on each metric are highlighted in bold text.

substantive efforts and assumptions. Especially, when the underlying physical process to construct albedo/emissivity from TSMP and tracing the atmospheric effects with satellite and solar geometry is very complex.

As mention in Section 5.2.3, we propose to use channel attention to determine the relative importance of TSMP input variables. Implementation details about the module are provided in Section 5.2.3 and Fig. 5.4. We used channel attention directly before the patch embedding for the U-Net model. To disentangle the correlation between NDVI and BT, we trained two separated models. One to predict NDVI and another one to predict BT. Note that we only used channel attention for this experiment. Fig. 5.16 provides example attentions induced for each input variable from COSMO, CLM, and ParFlow with respect to all weeks in the test and validation sets. The attention value is the mean value and it represents the variable importance to predict NDVI (BT). Error bars show how the attention changes across the weeks and input samples. We observe that the distributions of attention values for NDVI and BT is close. This indicates that the importance of highly relevant input variables are probably shared for both NDVI and BT. In addition, the standard deviations (error bars) suggest that the choice of prior explanatory variables is not trivial since the relative importance can change with time and input samples.

Overall, not all variables are relevant for the model. For COSMO, atmosphere water divergence (hudiv), humidity-related variables (hus, hur), precipitation variables (pr, prc, prg), surface air pressure (ps), drag coefficient of heat (tch) and geo-potential height (zg200) receive the highest attention from the DL model. For CLM, all variables are considered important with snowfall flux (prsn) and precipitation on ground (prso) being less important. Regarding ParFlow variables, it can be seen that the model considers most underground water-related variables as relatively important. This is intuitive since water and the amount of underground water storage are important factors for the vegetation growth. The availability of groundwater supply can reduce vulnerability to agricultural drought (*Meza et al.*, 2020; *Ma et al.*, 2021). Some previous studies showed that precipitation and tempera-

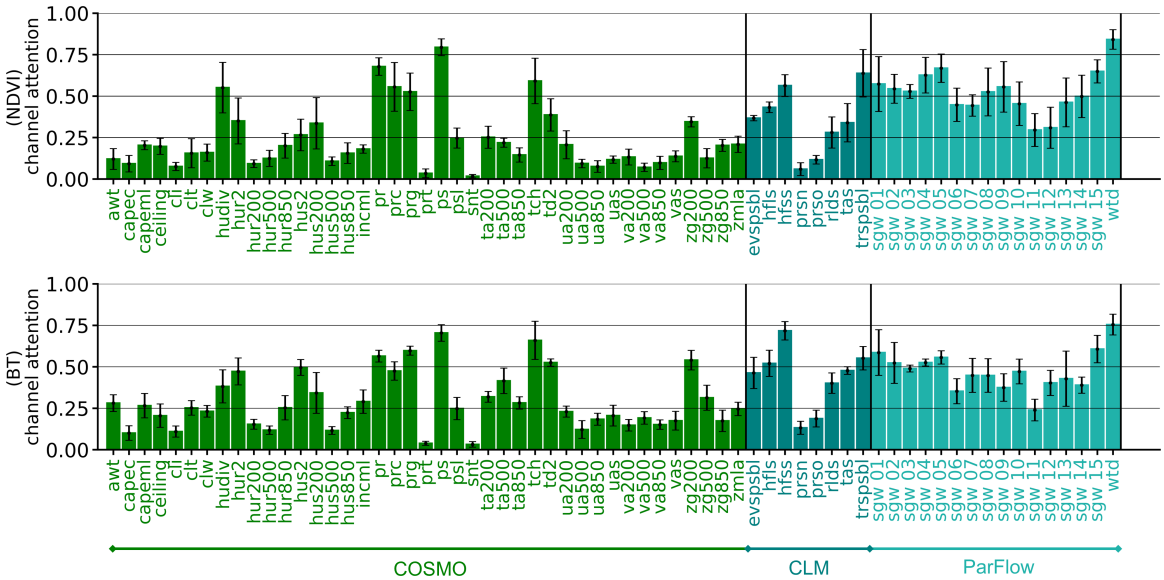


Figure 5.16: Channel attention for TSMP input variables. The activations are shown for both NDVI (top) and BT (bottom) with respect to all weeks in the validation and test sets.

ture are strong predictors of NDVI (*Miao et al.*, 2015; *Wu et al.*, 2020; *Gao et al.*, 2023a). In addition, the climatology of long-term NDVI is highly correlated with precipitation and the biome classification (*Yang et al.*, 2021b). The relatively high value for zg200 in BT prediction can be explained as the decrease in zg200 increases the likelihood of heatwave occurrence (*Miralles et al.*, 2019). The attention values for COSMO can be interpreted as *Nagol et al.* (2009) showed that scattering and absorption in the atmosphere affect the visible and near infrared radiance considerably. *Shi et al.* (2018) and *Geiss et al.* (2021) analyzed the influence of clouds related parametrization on visible and infrared satellite images and found that the accuracy is closely related to the cloud representation. A further study about the impact of surface and air pressure and water and ice clouds on visible and near-infrared bands can be found in *Baur et al.* (2023). It needs to be emphasized that the correlations shown in Fig. 5.16 must not be interpreted as a causal reasoning. One main reason is that data in Earth science are subject to complicated interactions and inherently inter-dependent. There may be hidden confounding variables that influence the explanatory variables as well as the evolution of the climate and vegetation variability. It is also worth noting that the learned variable importance by machine learning models is dependent on how the variables are represented in the training data (*Betancourt et al.*, 2022). Furthermore, some variables have larger biases than others since TSMP was run in a free mode simulation. This may drive the model to rely less on such variables even if they are considered important in scientific literature. The same thing applies to highly correlated variables where changing the model architecture may alter dependencies as well (*Betancourt et al.*, 2022).

5.3.5 Ablation Studies

5.3.5.1 VGG Perceptual Loss

In Table 5.5, we provide an additional analysis about the impact of the perceptual VGG loss described in Eq. (5.11). When adding a perceptual loss for training, we observe a consistent improvement for all metrics while residuals are slightly bigger for the test set. As shown in Figs. 5.17 and 5.18, adding the loss $\mathcal{L}_{VGG}$ reduces the blurring effect and increases variability.

		NDVI					BT (K)				
	Loss function	MAE(↓)	RMSE(↓)	R2(↑)	R_p(↑)	R_s(↑)	MAE(↓)	RMSE(↓)	R2(↑)	R_p(↑)	R_s(↑)
Val	$\mathcal{L}_{MAE}$	0.0274	0.0364	0.8744	0.9400	0.9139	1.9562	2.5945	0.9255	0.9664	0.9597
	$\mathcal{L}_{MAE} + \mathcal{L}_{VGG}$	**0.0270**	**0.0359**	**0.8781**	**0.9433**	**0.9184**	**1.8981**	**2.5433**	**0.9266**	**0.9679**	**0.9613**
Test	$\mathcal{L}_{MAE}$	**0.0266**	**0.0350**	**0.8819**	0.9443	0.9219	1.9642	2.6329	0.9181	0.9639	0.9610
	$\mathcal{L}_{MAE} + \mathcal{L}_{VGG}$	0.0268	0.0353	0.8795	**0.9452**	**0.9243**	**1.8730**	**2.5277**	**0.9227**	**0.9672**	**0.9642**

Table 5.5: Ablation study on the perceptual VGG loss described in Eq. (5.11). The metrics are shown for the validation and test sets as one set. The used model is a U-Net based on focal modulation. The best results on each metric are highlighted in bold text.

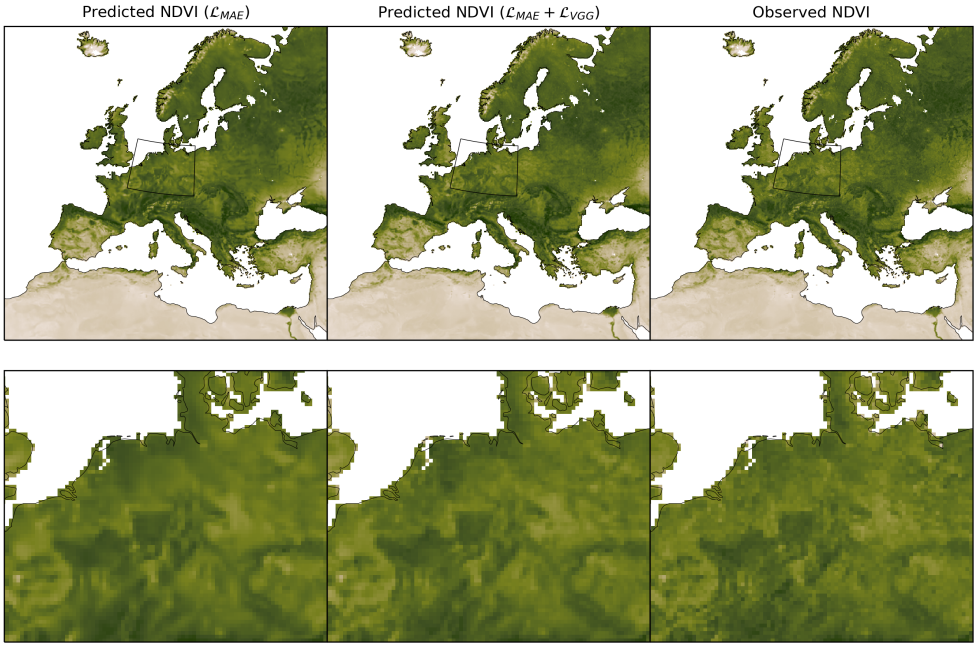


Figure 5.17: Impact of the perceptual VGG loss on NDVI predictions and image sharpness. The shown example is for the week 30 in the year 2018. Best seen in digital formats with colors.

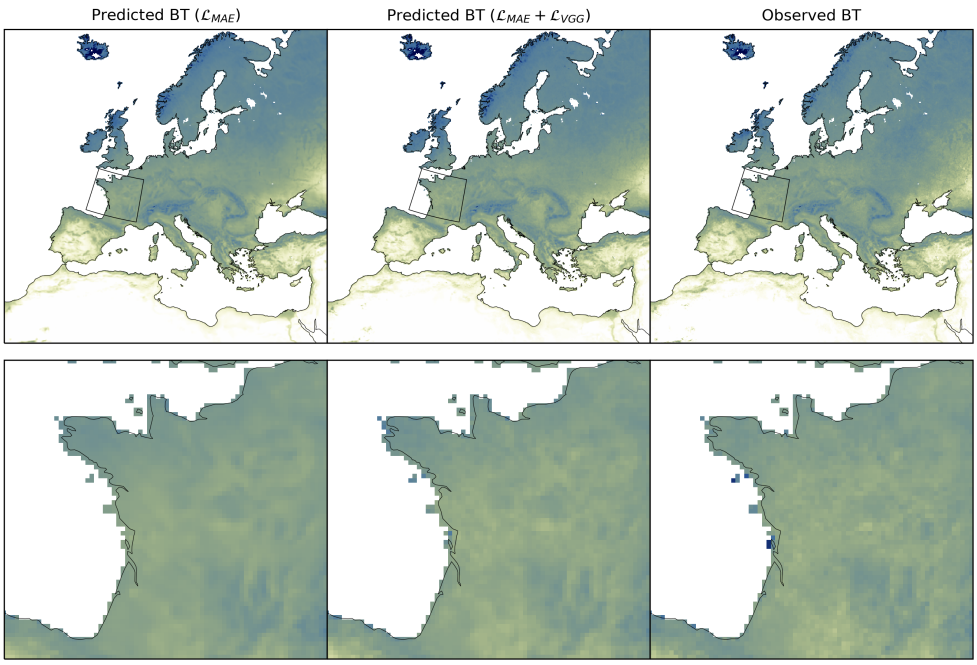


Figure 5.18: Impact of the perceptual VGG loss on BT predictions and image sharpness. The shown example is for the week 30 in the year 2018. Best seen in digital formats with colors.

5.3.5.2 Patch Embedding

Patch embedding with a patch size > 1 is commonly used in vision transformer architectures. The main aim of this embedding is to increase the channel dimension and reduce the computational demands of the self-attention modules. This can be done by merging and embedding neighborhood pixels/tokens thus reducing the spatial or temporal resolution. In Table 5.6, we show that decreasing the spatial dimension of the raw input for the encoder has negative effects on our image-to-image regression task in both quantitative and qualitative terms. This can be understood as the information was lost and the model struggles to output the original resolution. Note that for all experiments we keep using the down- and up-sampling with a factor of 2 in both encoder and decoder while we only change the patch size before the first encoder layer. To match the original spatial resolution, we used an additional bilinear up-sampling after the last decoder layer.

		NDVI					BT (K)				
	Patch size	MAE(↓)	RMSE(↓)	R2(↑)	R_p(↑)	R_s(↑)	MAE(↓)	RMSE(↓)	R2(↑)	R_p(↑)	R_s(↑)
Val	1×1	**0.0270**	**0.0359**	**0.8781**	**0.9433**	**0.9184**	**1.8981**	**2.5433**	**0.9266**	**0.9679**	**0.9613**
	2×2	0.0280	0.0369	0.8707	0.9374	0.9116	1.9372	2.6108	0.9243	0.9664	0.9604
	4×4	0.0291	0.0383	0.8625	0.9345	0.9075	2.0033	2.6957	0.9184	0.9633	0.9570
Test	1×1	**0.0268**	**0.0353**	**0.8795**	**0.9452**	**0.9243**	**1.8730**	**2.5277**	**0.9227**	**0.9672**	**0.9642**
	2×2	0.0271	0.0355	0.8786	0.9422	0.9185	1.9638	2.6669	0.9157	0.9645	0.9618
	4×4	0.0286	0.0375	0.8644	0.9363	0.9141	2.1741	2.9132	0.8977	0.9594	0.9580

Table 5.6: Impact of patch size for patch embedding before the first encoder layer. The metrics are shown for the validation and test sets. The used model is a U-Net based on focal modulation model. The best results on each metric are highlighted in bold text.

5.4 Summary

In this chapter, we presented a new deep learning approach for vegetation health prediction from a regional climate simulation. The developed model enabled the prediction of variables which are not part of the input simulation. In particular, we developed a vision transformer model with focal modulation to predict NDVI and BT images from a long-term TSMP (Ground to Atmosphere G2A) simulation at 0.11° resolution and on a weekly basis. We further validated the approach with NOAA remote sensing satellite observations and identified regions of uncertainty in the model predictions. As part of this, agricultural drought assessment was performed based on vegetation health products, namely VCI, TCI and VHI, which were derived from the predicted NDVI and BT, as well as long-term climatology. In this regard, the applicability of the model was spatially and temporally analyzed at a continental scale. Additionally, we extended the commonly used explanatory variables by using plenty of TSMP variables and analyzed their relative importance for the task with channel attention as an explainable AI method. The evaluation confirms that a DL model that was trained on observations has the capacity to predict NDVI and BT from a TSMP climate simulation with a sufficiently good agreement with real-world satellite observations.

CHAPTER 6

Identifying Spatio-Temporal Drivers of Extreme Events

This chapter is based on the following publication (*Shams Eddin and Gall*, 2024c):

Identifying Spatio-Temporal Drivers of Extreme Events
Mohamad Hakam Shams Eddin, and Juergen Gall
Advances in Neural Information Processing Systems 38 (NeurIPS), 2024
Doi: 10.48550/arXiv.2410.24075
Url: https://hakamshams.github.io/IDE

Below we summarize the contributions of each author.

- **Mohamad Hakam Shams Eddin**
  Hakam designed the method to identify anomalies and the framework to generate synthetic datasets. He also proposed to use the remote sensing and reanalysis data. Hakam conducted the experiments, preprocessed the data, implemented the codebase, and wrote the first draft of the paper.

- **Juergen Gall**
  Juergen proposed the idea of identifying the relations between anomalous events and droughts. He also proposed the idea of generating synthetic data. In addition to his financial support, Juergen contributed substantially through discussions and writing.

Contents

6.1 Introduction . . . 108
6.2 Method . . . 109
6.3 Synthetic data . . . 112
6.4 Experimental results and analysis . . . 115
6.4.1 Experimental Setup . . . 115
6.4.2 Generating Synthetic Data . . . 119
6.4.3 Experiments on Synthetic Data . . . 129
6.4.4 Ablation Studies . . . 133
6.4.5 Experiments on Real-world Data . . . 138
6.5 Summary . . . 151

In Focal-TSMP (Chapter 1.3.2), we addressed forecasting vegetation health and extreme agricultural droughts. The next step is to identify the drivers of extreme agricultural droughts. In this chapter, we propose a first approach and benchmarks to tackle the challenge of identifying spatio-temporal relations of impacts of extreme events and their drivers in climate data. These relations are not fully understood and there is a need of machine learning approaches to identify such spatio-temporal relations from data. The task, however, is very challenging since there are time delays between extremes and their drivers, and the spatial response of such drivers is inhomogeneous. In this chapter, we present an approach that is trained end-to-end to predict spatio-temporally extremes and spatio-temporally drivers in the physical input variables jointly. By enforcing the network to predict extremes from spatio-temporal binary masks of identified drivers, the network successfully identifies drivers that are correlated with extremes. We evaluate our approach on three newly created synthetic benchmarks, where two of them are based on remote sensing or reanalysis climate data, and on two real-world reanalysis datasets.

6.1 Introduction

A frontier research challenge is to understand the effects of global change on the magnitude and probability of extreme weather events (*Herring et al.*, 2022). Overall, the evolution of extreme events such as agricultural droughts results from stochastic processes (*Sillmann et al.*, 2017), conditions at ecosystem scales (*Ek and Holtslag*, 2004; *Miralles et al.*, 2019), and the interaction between the Earth land and atmospheric variables as a part of a complex system of feedbacks (*Geirinhas et al.*, 2023). However, the relative impacts of these factors differ depending on the event (*Sillmann et al.*, 2017). The time delays between extremes and their drivers vary seasonally (*Huang et al.*, 2015; *Wang et al.*, 2022a; *Dai et al.*, 2022; *Cao et al.*, 2022), and the spatial response of these drivers is inhomogeneous (*Miralles et al.*, 2019). A major challenge is therefore to model the spatio-temporal relations between extremes and their drivers during the development of these events (*Miralles et al.*, 2019). The overarching goal of this modelling is to improve our understanding of the patterns and impacts of such events. This would improve our ability to project duration and intensity of extreme events and hence assisting in adaptation planning (*Stott et al.*, 2016; *Hao et al.*, 2018).

In this chapter, we propose an approach that identifies spatio-temporal drivers in multivariate climate data that are correlated with the impact of extreme events. For the extreme events, we focus on agricultural droughts as an example, which can be measured by extremely low values of the vegetation health index (VHI). As drivers for such measurable extreme events, we consider anomalies in atmospheric and hydrological state variables like temperature or soil moisture, as well as land-atmosphere fluxes like evaporation. The task of identifying end-to-end spatio-temporal drivers for measurable impacts of extremes has not been addressed before, and it is very challenging since the drivers can occur earlier in time and at a different location than the measured extreme event as illustrated in Fig. 6.1.

To address this challenging task, we propose a network that is trained to predict spatio-temporally extremes. Instead of simply predicting the extremes, the network quantizes the spatio-temporal input variables into binary states and predicts the extremes only from the spatio-temporal binary maps for each time series of input variables. In this way, the network is enforced to identify only drivers in the input variables that are spatio-temporally correlated with extreme events. While the network is

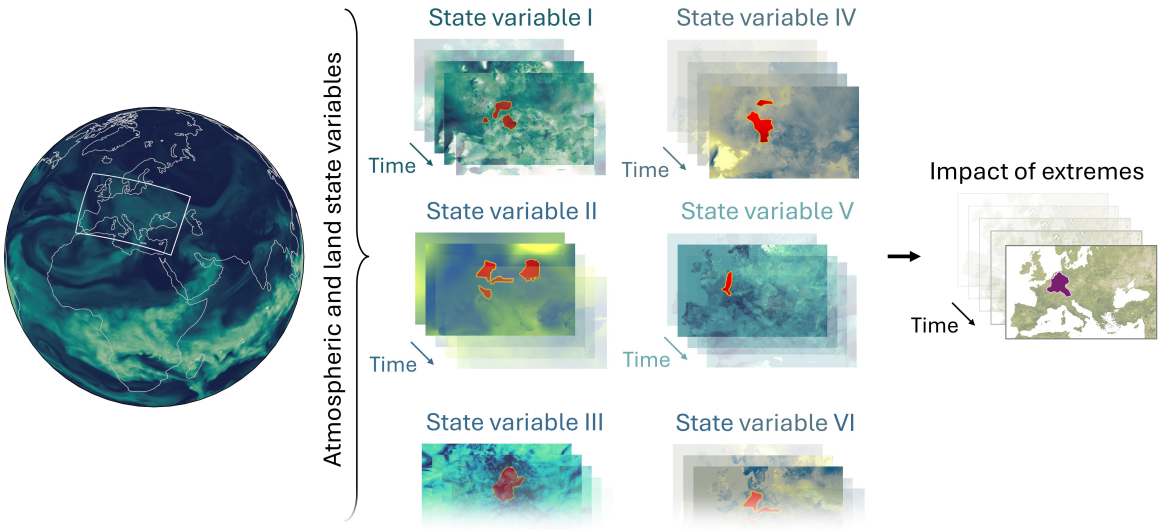


Figure 6.1: Overview of the objective of this work. We are interested in identifying spatio-temporal relations between the measurable impacts of extremes like the vegetation health index ■ and their drivers ■. As drivers, we focus on anomalies in state variables of the land-atmosphere and hydrological cycle. The task is very challenging since the drivers can occur at a different region than the extreme event and earlier in time.

trained using annotations of impacts of extreme events, which can be derived from remote sensing or reported data, we do not have any annotations of drivers or anomalies in the input variables.

Since drivers of extreme events are not fully understood, we cannot quantitatively measure the accuracy of the identified drivers on real-world data. We therefore propose a framework for generating synthetic data that can be used to assess the performance of our model as well as other baselines quantitatively. We evaluate our approach on three synthetic datasets where two of them are based on remote sensing or reanalysis climate data. Our evaluation shows that our approach outperforms approaches for interpretable forecasting, spatio-temporal anomaly detection, out-of-distribution detection, and multiple instance learning. Furthermore, we conduct empirical studies on two real-world reanalysis climate data.

The rest of this chapter is organized as follows. The methodology is described in Section 6.2. A framework to generate synthetic data is given in Section 6.3. Experimental results and analyses are given in Section 6.4. Finally, a summary is provided in Section 6.5.

6.2 Method

Our aim is to design a model that is capable of identifying spatio-temporal drivers of extremes in multivariate climate data, e.g., Earth observations or climate reanalysis. In particular, we want to identify anomalies that are spatio-temporally related to extreme events like agricultural droughts. This is different to standard anomaly detection since we are not interested in all anomalies, but in spatio-temporal configurations of variables that potentially cause an extreme event with some time delay and at a potentially different location as illustrated in Fig. 6.1. To achieve this, we propose a network that is trained end-to-end on observed extremes where we focus on agricultural droughts.

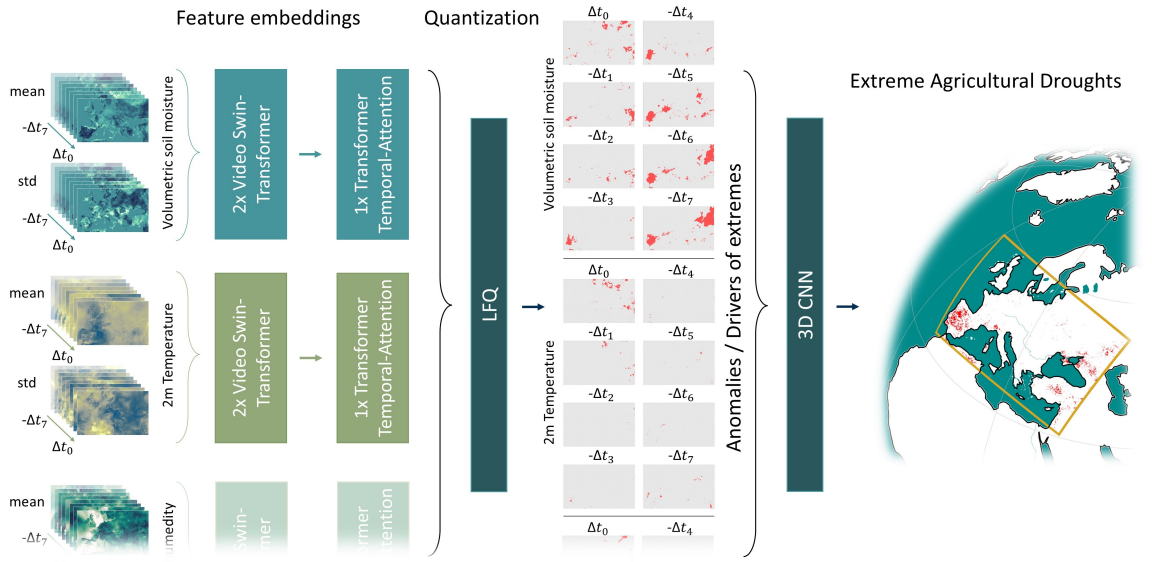


Figure 6.2: An overview of the proposed model to identify the spatio-temporal relations between extreme agricultural droughts and their drivers. The input variables are first encoded into features. In a subsequent step, a lockup free quantization layer (LFQ) takes the extracted features and classifies the variables into a binary representation of drivers, where we consider the drivers as anomalous events in the input variables. Finally, a classifier is used to predict impacts of extreme events from the identified drivers.

The network classifies the input variables before an extreme event occurs into spatio-temporal drivers without additional annotations besides the annotated extremes. The network then aims to predict future extremes based on the identified drivers. It needs to be noted that we are interested in input variables that do not define an extreme event, but we aim to find anomalies in input variables that are correlated with the occurrence of an extreme event. We will thus denote potential drivers as anomalies.

An overview of our approach is shown in Fig. 6.2. The input are weekly climate variables at sequential time steps $(\Delta t_{-7}, \ldots, \Delta t_0)$ and the model is composed of three main parts. First, a feature extractor extracts relevant features from each input variable independently. The second component is a quantization layer that takes the extracted features as input and classifies the input variables into drivers. The role of the quantization layer is to transform the input variable into a binary representation (1 = drivers and 0 otherwise). This ensures that no additional information is encoded as an input to the subsequent classifier except that if the input variable at a specific location and time is a driver or not. The third component is a classifier that takes as input the variables, location, and time where drivers have been identified and predicts where extreme droughts occur at the time step Δt_0. All model components are trained jointly.

We denote the input data as $\mathbf{X} \in \mathbb{R}^{V \times C \times T \times Lat \times Lon}$, where V is the input variables, C is the channel dimension for each variable (e.g., mean and standard deviation of the week), T is the temporal resolution $(\Delta t_{-T+1}, \ldots, \Delta t_0)$, and Lat and Lon are the spatial extensions. The model has two outputs, $\mathbf{Q} \in \mathbb{Z}_2^{V \times T \times Lat \times Lon}$ representing the binary classification of the input variables into potential drivers of the extreme events, and probabilities $\mathbf{E} \in \mathbb{R}^{Lat \times Lon}$ to predict extreme events at

the time step Δt_0. In this work, we formulate this as follows:

$$(g_\vartheta \circ q_\varphi \circ f_\theta) : (\{\mathbf{X}^{\Delta t_{-T+1}}, \ldots, \mathbf{X}^{\Delta t_0}\}) \rightarrow (\{\mathbf{Q}^{\Delta t_{-T+1}}, \ldots, \mathbf{Q}^{\Delta t_0}\}; \mathbf{E}^{\Delta t_0}) \,, \tag{6.1}$$

where g_ϑ, q_φ, and f_θ are the mapping functions parameterized as neural networks. In the following, we describe the model components:

Feature extraction. First, embedded features $f_\theta : \mathbf{X} \rightarrow \mathbf{Z} \in \mathbb{R}^{V \times K \times T \times Lat \times Lon}$ are extracted independently for each input variable $v \in V$ with K embedding dimensions and parameters θ. The rationale behind this independence is to prevent that drivers leak into other variables. We use the Video Swin Transformer model (*Liu et al.*, 2022b) as backbone to capture long-range interrelations across time and space. We also run experiment with state space model Mamba (*Gu and Dao*, 2024) as a replacement for Swin Transformer. The input $\mathbf{X}$ is projected into a higher feature dimension K and followed by two Video Swin Transformer layers. The first layer has two consecutive 3D shifted window blocks for a spatio-temporal feature extraction. The second layer consists of one block for a temporal feature extraction. The later is useful to focus only on the temporal evolution of the variables. An ablation study regarding the backbone is provided in Section 6.4.4.1.

Quantization layer. The role of this layer is to map $\mathbf{Z}$ from an embedded space into a compact binary representation $\mathbf{Q}$ suitable for detecting drivers. Using vector quantization (VQ) (*Van Den Oord et al.*, 2017), each embedded feature vector $z \in \mathbf{Z}$ is assigned into a learnable codebook feature vector $z_q \in \mathbf{Z}_q$ based on the Euclidean distance:

$$VQ : z \rightarrow z_q, \text{where } q = \underset{q \in \{1,\ldots,Q\}}{\arg\min} \|z - z_q\|_2 \,, \tag{6.2}$$

where Q is the size of the codebook. Recently, lookup-free quantization (LFQ) (*Yu et al.*, 2024) substitutes the learnable codebook with a set of integers $\mathbb{Q}$ with $|\mathbb{Q}| = Q$ and represents the embedding space as a Cartesian product of binary numbers. This omits the need for a distance metric to do the nearest vector assignment and simplifies the quantization. Based on experimental results, we built the vector quantizer on LFQ with two integers $Q = 2$ ($q = 1$ for drivers and $q = 0$ otherwise). Given a feature vector z, LFQ first maps z into a scalar value $z_l \in \mathbf{Z}_l \subset \mathbb{R}^{V \times 1 \times T \times Lat \times Lon}$. For multi-modality, we use two sequential 3D CNNs on each input variable followed by a shared linear layer that maps z to z_l and reduces the dimensions from K to 1. The quantization is then given by the sign of z_l:

$$z_q = \text{Linear}\big(\text{sign}(z_l)\big) = \text{Linear}(-\mathbb{1}_{\{z_l \leqslant 0\}} + \mathbb{1}_{\{z_l > 0\}}), \;\; \text{and} \;\; q = \mathbb{1}_{\{z_l > 0\}} \,, \tag{6.3}$$

where $q \in \mathbf{Q}$ represents the class ($q = 1$ or $q = 0$), and Linear is a linear layer that converts $\text{sign}(z_l)$ back to the dimension K of the input after the quantization. Note that $\mathbf{Z}_q$ has only two unique vectors $z_{q=1}$ for a driver and $z_{q=0}$ otherwise.

Prediction of extreme events. We use a classifier that predicts the probably of extreme events $\mathbf{E}$ at the time step Δt_0 from the identified drivers $\mathbf{Z}_q$. We use a 3D CNN classifier instead of a transformer to reduce the computations. For training, we only know the ground truth of extremes at time step Δt_0 denoted by $\hat{\mathbf{E}} \in \mathbb{Z}_2^{Lat \times Lon}$. While we could compute the cross-entropy between $\mathbf{E}$

and $\hat{\mathbf{E}}$, we found that a single 3D CNN is insufficient to detect all drivers that are correlated with an extreme event. Instead, we use $V+1$ 3D CNNs where each predicts $\mathbf{E}_v$. While the first V 3D CNNs take the identified drivers for a single variable v as input, the last one takes the identified drivers of all variables as input. The multiple CNNs are only used for training. During inference, $\mathbf{E}$ is only predicted by the multivariate CNN where all variables are jointly used. The loss is thus given by

$$\mathcal{L}_{(extreme)} = -\sum_{v=1}^{V+1} \big(\hat{\mathbf{E}} \log(\mathbf{E}_v) + (1-\hat{\mathbf{E}}) \log(1-\mathbf{E}_v)\big)\mathbf{S}\,, \tag{6.4}$$

where $\mathbf{S} \in \mathbb{Z}_2^{T \times Lat \times Lon}$ is a mask for the valid regions. We also utilize a weighted version of $\mathcal{L}_{(extreme)}$ to mitigate the class imbalance issue (Section 6.4.4.3). While the loss $\mathcal{L}_{(extreme)}$ ensures that extreme events can be predicted from the identified drivers, we need to add standard loss terms for the quantization to ensure that the learned codes and thus drivers are compact:

$$\mathcal{L}_{(quantize)} = \lambda_{(commit)} \|\mathbf{Z}_l - \mathrm{sg}(\mathrm{sign}(\mathbf{Z}_l))\|_2^2 + \lambda_{(ent)} \mathbb{E}[H\big(\mathrm{sign}(\mathbf{Z}_l)\big)] - \lambda_{(div)} H[\mathbb{E}\big(\mathrm{sign}(\mathbf{Z}_l)\big)]\,. \tag{6.5}$$

The commitment loss $\|\mathbf{Z}_l - \mathrm{sg}(\mathrm{sign}(\mathbf{Z}_l))\|_2^2$ prevents the outputs of the encoder from growing and encourages $\mathbf{Z}_l$ to commit to the codes (*Van Den Oord et al.*, 2017), where sg stands for the stopgradient operator with zero partial derivative. The term $\mathbb{E}[H\big(\mathrm{sign}(\mathbf{Z}_l)\big)]$ encourages that the entropy per quantized code is low (*Yu et al.*, 2024; *Jansen et al.*, 2020), meaning that it provides more confident assignments. Whereas the term $H[\mathbb{E}\big(\mathrm{sign}(\mathbf{Z}_l)\big)]$ increases the entropy inside the batch to encourage the utilization of all codes (*Yu et al.*, 2024; *Jansen et al.*, 2020). The last important ingredient is a loss that ensures that only spatio-temporal regions are identified that correlate with an extreme event. To this end, we look at regions and intervals where no extreme event occurred and use these examples without drivers. Formally, we use $\hat{\mathbf{E}}_t \in \mathbb{Z}_2^{Lat \times Lon}$ as the union of extreme ground truth at all time steps $(\Delta t_{-T+1}, \ldots, \Delta t_0)$ and compute the loss by

$$\mathcal{L}_{(driver)} = \lambda_{(driver)} |\mathbf{Z}_q - \mathrm{sg}(z_{q=0})| (1 - \hat{\mathbf{E}}_t)\mathbf{S}\,, \tag{6.6}$$

where $z_{q=0}$ is the quantization code for normal data without drivers. The model is trained end-to-end with the joint optimization of the loss function:

$$\min_{\theta,\varphi,\vartheta} \underbrace{\mathcal{L}_{(extreme)}\big(\mathbf{E}, \hat{\mathbf{E}}, \mathbf{S}\big)}_{\text{predicts extremes from drivers}} + \underbrace{\mathcal{L}_{(quantize)}\big(\mathbf{Z}_l\big)}_{\text{encourages confident quantization}} + \underbrace{\mathcal{L}_{(driver)}\big(\mathbf{Z}_q, \hat{\mathbf{E}}_t, \mathbf{S}, \mathbf{Z}_{q=0}\big)}_{\text{assigns drivers to the same code in the codebook}}\,, \tag{6.7}$$

where $\theta, \varphi, \vartheta$ are the learnable parameters for the feature extractor, quantizer, and extreme classifier, respectively, and $\lambda_{(commit)}$, $\lambda_{(ent)}$, $\lambda_{(div)}$, and $\lambda_{(driver)}$ are weighting parameters. Ablation studies are provided in Section 6.4.4.

6.3 Synthetic data

Although ground truth for extreme droughts can be obtained from remote sensing, an important methodological question remains as how to reliably have a meaningful quantitative evaluation of the identified drivers and their relations to the extreme events. To the best of our knowledge, ground truth

labels of drivers or anomalies which are correlated with extreme events impacts barely exist. This is especially the case when it comes to drivers or anomalous events which are related to a specific definition of an extreme event within the Earth system (e.g., drought or flood). To solve this critical issue, we designed a framework to generate artificial datasets which can be adopted to the task. The framework generates data that mimics the properties of Earth observations including drivers and anomalies. We are aware that the dynamic of the synthetic data are simplified compared to real Earth observations. However, we rely on this generated dataset to perform the quantitative evaluation of the proposed approach. Our framework is inspired by and related to *Flach et al.* (2017). However, we differ in the following two aspects: (a) First, we aim to generate multivariate anomalous that are correlated with a specific extreme event, while their aim is to generate anomalous that can occur simultaneously in multivariate data streams similar to *Xu et al.* (2022), *Lee et al.* (2021), *Zhong et al.* (2025), and *Tuli et al.* (2022). Think of an increasing/decreasing of temperature, existing approaches are interested in temperature anomalies regardless of the subsequent extreme events they might cause, while we focus on temperature anomalies that can cause a particular extreme events in the near future. (b) Second, we generate the synthetic data based on real-world data stream signals, while they use trigonometric functions (e.g., sine function) to mimic Earth observations across spacetime.

In the following, we explain our overall framework in more details:

1. First, we generate the normal base signals $\mathbf{\Psi} \in \mathbb{R}^{V \times T \times Lat \times Lon}$ for a set of different variables V, where T is the temporal extension, and Lat and Lon are the spatial extensions. For instance to generate synthetic data of 2m temperature from CERRA reanalysis, the normal signal at a specific time and location is generated based on the typical value of 2m temperature at that time and location (e.g., the mean or median value from a long-term climatology). We synthesize CERRA (Section 2.2.5) $\mathbf{\Psi}$ signals by taking the mean values from the CERRA climatologoy pixel-wise. This represents the typical value of $\mathbf{\Psi}$ at specific time (week) and location (lat, lon). By definition, $\mathbf{\Psi}$ inherits the intrinsic properties of the simulated variables including the existence of seasonality and correlations among variables.

2. In the next step, we generate anomalies conditioned on the occurrence of extremes. To achieve this, we induce binary extreme events $\mathbf{E}^{ex} \in \mathbb{Z}_2^{T \times Lat \times Lon}$ within the datacube and track the precise spatio-temporal location of these events. Similar to *Flach et al.* (2017), the type and

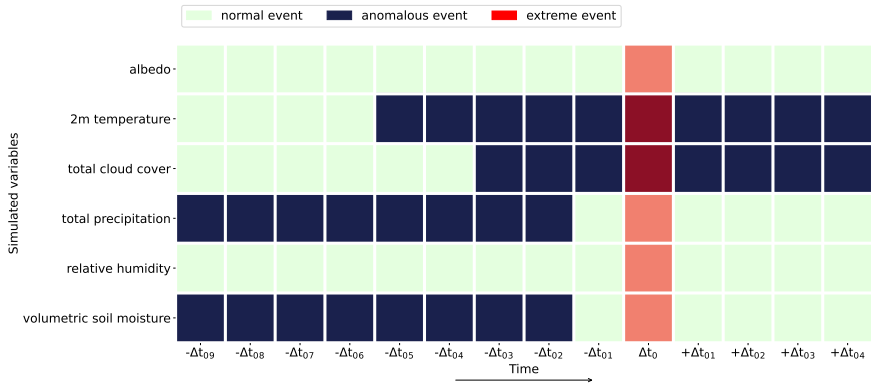


Figure 6.3: An example of the randomly generated coupling matrix between the synthetic variables and extremes for the synthetic CERRA dataset.

duration of extreme events vary within the datacube. For instance, we alter the duration of the events between long and short extreme events. The spatial distribution of the extreme event vary also between a local event at one pixel (LocalEvent), a rectangular event (CubeEvent), a Gaussian shape (GaussianEvent), an onset event that starts at specific time and lasts until the end of the series (OnsetEvent), and a random walk event that starts at a specific pixel and affecting neighboring pixels with time (RandomWalkEvent).

3. We randomly define a coupling matrix $\mathbf{M}$ between the variables and the extreme event. An example is shown in Fig. 6.3, e.g., the anomalous events for 2m temperature start 5 time steps ($-\Delta T_a$) before the extreme and last for 4 time steps afterwards ($+\Delta T_a$). While albedo and relative humidity are not coupled with the extremes. Based on $\mathbf{M}$ and $\mathbf{E}^{ex}$, we generate the binary anomalous events matrix $\mathbf{E}^a \in \mathbb{Z}_2^{T\times Lat\times Lon}$. Note that we generate anomalous events only for the variables that are defined to be correlated with the extremes. We consider these anomalies as the drivers for the extreme events.

4. Similar to steps (2)-(3), we generate binary random anomalous events $\mathbf{E}^r \in \mathbb{Z}_2^{T\times Lat\times Lon}$. However, these events are uncorrelated with the extremes and generated randomly for all variables.

5. We sample noise signals $\mathbf{N} \in \mathbb{R}^{V\times T\times Lat\times Lon}$ for each variable. The noise signals could be sampled from a normal Gaussian distribution (GaussianNoise), a standard Cauchy distribution (CauchyNoise), a double exponential distribution (LaplaceNoise), or from a spatio-temporal correlated noise across the datacube (RedNoise).

6. Using the indices $v \in \{1,\dots,V\}$, $t \in \{1,\dots,T\}$, $lat \in \{1,\dots,Lat\}$, and $lon \in \{1,\dots,Lon\}$, the synthetic signal $\mathbf{\Phi}_{(v,t,lat,lon)}$ is generated with the following formula:

$$\mathbf{\Phi}_{(v,t,lat,lon)} = \mathbf{\Psi}_{(v,t,lat,lon)} + \mathbf{\Lambda}_{(v,t,lat,lon)}\cdot\mathbf{\Theta}_{(v,t,lat,lon)}\,, \tag{6.8}$$

$$\begin{aligned}\mathbf{\Theta}_{(v,t,lat,lon)} = {} & \mathbf{\Psi}_{(v,t,lat,lon)}\cdot(2^{(kb.(\mathbf{E}^r_{(v,t,lat,lon)}\vee\mathbf{E}^a_{(v,t,lat,lon)}))}-1)\\ & + \mathbf{N}_{(v,t,lat,lon)}\cdot 2^{(kn.(\mathbf{E}^r_{(v,t,lat,lon)}\vee\mathbf{E}^a_{(v,t,lat,lon)}))}\\ & + ks.(\mathbf{E}^r_{(v,t,lat,lon)}\vee\mathbf{E}^a_{(v,t,lat,lon)})\cdot\sigma_N\,,\end{aligned} \tag{6.9}$$

$$\mathbf{\Lambda}_{(v,t,lat,lon)} = \begin{cases}\dfrac{\mathbf{\Delta}_{(v,t,lat,lon)}}{\delta}, & \text{if } \mathbf{E}^a_{(v,t,lat,lon)}=1,\\ +1, & \text{otherwise.}\end{cases}, \tag{6.10}$$

$$\mathbf{\Delta}_{(v,t,lat,lon)} = \begin{cases}-1, & \text{if } \mathbf{\Theta}_{(v,t,lat,lon)}\leqslant 0,\\ +1, & \text{otherwise,}\end{cases}, \tag{6.11}$$

where $\mathbf{\Theta}_{(v,t,lat,lon)}$ is the induced anomaly, kb, kn, and ks are control parameters for the events magnitudes, σ_N is the standard deviation of the noise signal, $\delta \in \{-1,1\}$ is the predefined coupling sign with extremes from $\mathbf{M}$, and $\mathbf{\Lambda}_{(v,t,lat,lon)}$ controls the sign of the induced anomaly for extremes (e.g., a deficiency in soil moisture ($\delta=-1$) and an increased in temperature ($\delta=+1$) during an extreme drought event).

Using this framework, we generated 3 types of datasets; (a) Synthetic CERRA reanalysis, (b) Synthetic NOAA remote sensing , and (3) Synthetic artificial data. We generate the NOAA base

signals from *Yang et al.* (2020). For the NOAA and artificial datasets, we also considered artificial linear (LinearCoupling) and non-linear (QuadraticCoupling) dependencies among the variables as additional data properties. For instance to generate a new dependent base signal from independent bases:

$$\mathbf{\Psi}_{(t,lat,lon)} = \begin{cases} \sum_{v=1}^{V} w_{(v)} . \Psi_{(v)}, & \text{if LinearCoupling,} \\ \sum_{v=1}^{V} \frac{1}{\sqrt{2}} w_{(v)} . (\Psi_{(v)}^2 - 1), & \text{otherwise,} \end{cases} , \tag{6.12}$$

$$\mathbf{E}^r_{(t,lat,lon)} = \vee_{v=1}^{V} \mathbf{E}^r_{(v,t,lat,lon)} , \tag{6.13}$$

where $w_{(v)} \in \mathbb{R}^V$ are weighted coefficients sampled either from a normal (NormWeight) or a Lablace (LaplaceWeight) distribution. In addition, we add an option to generate disturbed weights where weights vary spatio-temporally based on the locations of anomalies. Anomalous events that occur in one of the independent base signals propagate to the new generated dependent signals. The third dataset (Artificial) does not depend on real-world data but rather consists of basis signals with trigonometric functions where we also add a linear latitudinal gradient (latGrad). Finally, we mask out some regions to exhibit no anomalies, e.g., pixels over water surface. Examples of the generated synthesized dataset can be found in Section 6.4.2

6.4 Experimental results and analysis

First, we describe the experimental setting in Section 6.4.1. Then we describe the generated synthetic data for evaluation in Section 6.4.2. In Section 6.4.3, we conducted experiments on the generated synthetic datasets and empirically verified the effectiveness of the proposed design compared to the baselines. Section 6.4.4 includes the ablation studies. Finally, we validate the model on two real-world datasets over five continents in Section 6.4.5.

We use Swin Transformer backbone in all experiments if not otherwise stated.

6.4.1 Experimental Setup

6.4.1.1 Datasets

Defining extreme agricultural droughts. We are interested in a specific impact of extreme events namely extreme agricultural drought. To define such extreme event, we rely on the observational satellite-based vegetation health index (VHI) obtained from NOAA (Section 2.2.3). This remote sensing product cannot be directly derived from the input reanalysis, which makes the task very challenging. VHI approximates the vegetation state based on a combination of the brightness temperature and normalized difference vegetation index (VHI $= 0$ for unfavorable condition and VHI $= 100$ for favorable condition). Extreme agricultural droughts are usually defined as VHI < 26 (*Yang et al.*, 2020). The dataset has a temporal coverage of 1981-onward and is provided globally on a weekly basis. We mapped this dataset into the same domains of the reanalysis data and used this dataset as ground truth for extreme events. Note that VHI is a general vegetation index and should be interpreted carefully. Details about this vegetation index, the dataset and preprocessing are provided in Sections 2.2.3 and 2.2.3.3.

Reanalysis data. We conducted the experiments on two real-world reanalysis datasets; CERRA reanalysis (Section 2.2.5) and ERA5-Land (Section 2.2.4). The datasets are aggregated on a weekly basis within an overlapped period with the remote sensing data. We conduct the experiments on CERRA domain over Europe and 6 CORDEX domains (*Giorgi et al.*, 2009), separately. More details regarding the variables and the domains along with the training/validation/test splits are provided in the Sections 2.2.5 and 2.2.4.

Synthetic data. As mention in Section 6.3, we propose to generate synthetic to perform the quantitative evaluation.

Following the described framework in Section 6.3, we synthesized 3 datasets; (a) Synthetic CERRA reanalysis, (b) Synthetic NOAA remote sensing , and (3) Synthetic artificial data. Each generated dataset consists of 52×46 time steps corresponding to overall 46 years of simulated data in 7-day intervals; 34 years for training, 6 subsequent years for validation and the last 6 years for testing. The challenge is to identify the drivers, e.g., the anomalous events that are correlated with extreme events. The generated dataset are shown in Section 6.4.2.

6.4.1.2 Implementation Details

We set the hidden dimension K to 16 by default. The temporal resolution is $T = 6$ for the synthetic data and $T = 8$ for real-world data. Since, seasonal cycles are typical in climate data, we deseasonalize locally by subtracting the median seasonal cycle and normalizing by the seasonal variance for each pixel.

The training was done mainly on clusters with NVIDIA A100 80GB and NVIDIA A40 48GB GPUs. In the following, we highlight the main technical implementations and hyperparameters for training:

Synthetic data. For the synthetic data we set the embedding dimension $K = 16$. We use one layer Video Swin Transformer with {depth=[2, 1], heads=[2, 2], window size=[[2, 4, 4], [6, 1, 1]]}. For the quantization layer, we use two sequential 3D CNN layers on each variable with: {[kernel=(3, 3, 3), stride=(1, 1, 1), padding=(1, 1, 1)]} and followed by a shared linear layer. The classifier g_ψ consists of the following layers: {[3D CNN, kernel=(3, 3, 3), stride=(3, 1, 1), padding=(0, 1, 1)], [3D CNN, kernel=(2, 3, 3), stride=(2, 1, 1), padding=(0, 1, 1)]}. We set $\lambda_{(ent)} = \lambda_{(div)} = 0.1$, $\lambda_{(anomaly)} = 100$, and $\lambda_{(commit)} = 3$. The models were trained with Adam optimizer *Kingma and Ba* (2014) for 100 epochs with a batch size of 4. We use a linear warm up of 2 epochs and a cosine decay with an initial learning rate of 2×10^{-3} and a weight decay of 3×10^{-3}.

Reanalysis data. We set the embedding dimension $K = 24$ for CERRA and CAS-11 datasets and $K = 16$ for the rest of ERA5-Land datasets. We use one layer Video Swin Transformer with {depth=[2, 1], heads=[2, 2], window size=[[2, 4, 4], [8, 1, 1]]}. The quantization layer is similar to the one for the synthetic data. The classifier g_ψ consists of the 3 layers each has {[D CNN, kernel=(2, 3, 3), stride=(2, 1, 1), padding=(0, 1, 1)}. We set $\lambda_{(ent)} = \lambda_{(div)} = 0.1$, and $\lambda_{(anomaly)} = 100$ by defaults and $\lambda_{(commit)} = 1.0$ for CAS-11. For CERRA, we set $\lambda_{(ent)} = \lambda_{(div)} = 0.01$. Due to the high resolution on the reanalysis data, we use gradient checkpoint during training. For CERRA reanalysis, we also cut the boundaries between low and high latitudes focusing on the central region.

This results in a final grid with 512×832 cells. To handle missing data and temporal gaps in the input reanalysis data, we first normalize the data using the pre-computed statistics and then replace the invalid pixels with zero values.

Additional details and code are available at https://github.com/HakamShams/IDEE and https://github.com/HakamShams/Synthetic_Multivariate_Anomalies.

6.4.1.3 Baselines

We evaluate our approach with an interpretable forecasting approach using integrated gradients (*Sundararajan et al.*, 2017) and 8 baselines from 3 main related categories in anomaly detection; one-class unsupervised (*Schölkopf et al.*, 1999; *Liu et al.*, 2012, 2023b), reconstruction-based (*Astrid et al.*, 2021; *You et al.*, 2022), and multiple instance learning (*Sultani et al.*, 2018; *Wan et al.*, 2020; *Tian et al.*, 2021). Note that these baselines are not directly applicable to the task we address in this chapter. We modified and trained all baselines from scratch. For this, we relied on the officially released codes and started from the default hyperparameters. We also compare to a naive baseline. The implementation details of these baselines are given below:

Naive. This baseline labels all variables as drivers for pixels where extreme events occur.

Integrated Gradients (*Sundararajan et al.*, 2017). This is an axiomatic attribution method which can be used as a post-hoc method to explain the model prediction. We trained two models that predict extreme events directly from the input variables and then we applied a post-hoc integrated gradients from Pytorch Captum (*Kokhlikyan et al.*, 2020). Both models use the same Swin Transformer backbone as our model but without the anomaly detection step. For Integrated Gradients II, we added a cross attention. For these baselines, we compute the gradient only with respect to predicted extremes and computed a different threshold for each variable separately.

Isolated Forest (IF) (*Liu et al.*, 2012). This is an ensemble of random trees that isolate input instances by randomly selecting abounded splits for features. A shorter averaged path for a recursive partitioning implies an anomalous input. We empirically set the number of estimators to 100.

OCSVM (*Schölkopf et al.*, 1999). This is a one-class support vector machine solved using stochastic gradient descent. We use a radial basis kernel with 100 components.

For IF and OCSVM, we extracted multivariate features as the distances between the input variables and then train a model on each input variable separately. We sampled 400k normal data points (locations without extreme flags) for training and defined the expected ratio of anomalous to be roughly equivalent to the ratio of extreme events in the data.

SimpleNet (*Liu et al.*, 2023b). This method is categorized as an embedding-based one-class algorithm for anomaly detection. In SimpleNet, a feature extractor is first used to extract local features from normal data followed by a feature adaptor. Then an anomaly feature generator induces anomalous features by adding a Gaussian noise in the feature space. Finally, a discriminator is trained

to distinguish between the normal and anomalous generated features. During inference, the feature generator is discarded and the trained discriminator is expected to separate anomalous from normal input features. During inference, we compute the median anomaly score for both normal pixels and pixels with extreme flags. Then, we set a threshold for anomalous pixels based on the average of the two former median anomaly scores. We trained SimpleNet with our pretrained model as a feature extractor and set the feature dimension for feature adaptor to 512. The extracted features are scaled by 10^{-2} and the anomalous feature generator uses a Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma = 1.5)$. The discriminator composes of two linear layers with a hidden dimension of 96. We set th^+ and th^- to 1.0 and trained with AdamW optimizer for 48 epochs with a batch size of 2 and a learning rate of 3×10^{-4} with 1×10^{-5} weight decay.

STEALNet (*Astrid et al.*, 2021). Also called synthetic temporal anomaly guided end-to-end video anomaly detection. This is a reconstruction-based algorithm. STEALNet was trained to maximize the reconstruction loss for locations with extreme flags and to minimize the reconstruction loss otherwise. We set the dimensions for the auto-encoder as 96, 128, and 256. We trained with a batch size of 8 for 100 epochs using Adam optimizer with 1×10^{-3} weight decay and a learning rate of 3×10^{-4}.

UniAD (*You et al.*, 2022). This method is a unified model for multi-class anomaly detection. During training, a transformer model is trained to reconstruct masked tokens from the normal data. To prepare the input, we replaced pixels over extremes with their median values from the climatology. We set the hidden dimension to 96 with 3 heads and a depth of 3 for both the encoder and the decoder. The masked neighbourhood was set to 9×9. We down-sample the features by a factor of 2 and up-sample the output at the end of the network. The model was trained with AdamW for 64 epochs with a learning rate of 1×10^{-3} with 1×10^{-4} weight decay.

For STEALNet and UniAD, we compute during inference the mean reconstruction loss for both normal pixels and pixels with extreme flags. Then, we set a threshold for anomalous pixels based on the average of these two values.

DeepMIL (*Sultani et al.*, 2018). This is a multiple instance learning for anomaly detection in surveillance videos. In MIL, each video is represented as a bag of snippets (instances). We define positive instances as pixels with extreme event flags where the exact information (which variable is anomalous) within the positive instances is unknown. In the original implementation, the ranking is enforced only on the top instance with the highest anomaly score in each bag. We modified the ranking loss to top-k and set $K = 100$.

ARNet (*Wan et al.*, 2020). The method is also called anomaly regression net is another MIL-based approach. The ranking loss is based on a top-k binary cross entropy. In addition, there is a center loss to reduce the intra-class discriminative features of normal instance. We set $\alpha = 400$ and $\lambda_{center} = 20$.

RTFM (*Tian et al.*, 2021). Or robust temporal feature magnitude learning is a MIL-based algorithm which learns to distinguish between the normal and anomalous scores by selecting the top-k snippets with the highest feature magnitudes. We modified the multi-scale temporal network (MTN)

to capture the local and global spatial dependencies between pixels. We set the dimension for MTN to 32, $K = 100$, the margin to separate features to $\times 10^2$ and $\alpha = 1 \times 10^{-4}$.

We use the same Video Swin Transformer backbone as our model to train DeepMIL, ARNet, and RTFM. During training, we noticed that the ranking loss becomes biases toward one variable when the ranking is computed among all variables, so we computed a loss for each variable and then average the losses. RTFM only worked when we added a cross attention between the variables. We suppose this to be related to the feature magnitude learning. We trained with a batch size of 2 using Adam optimizer and a learning rate of 6×10^{-4} for 100 epochs with a weight decay of 1×10^{-3}. We also added an instance dropout of 0.5.

6.4.1.4 Evaluation Metrics

For evaluation, we use the F1-score, intersection over union (IoU), and overall accuracy on both classes (OA).

6.4.2 Generating Synthetic Data

We follow the framework described in Section 6.3 to generate 3 synthetic datasets. Technical details on the synthetic CERRA are explained in Table 6.1 and the details on synthetic NOAA and the artificial data are shown in Tables 6.2 and 6.3, respectively. For instance, the synthetic CERRA dataset mimics a set of variables ($V = 6$) using statistics from the real-world CERRA reanalysis. We artificially correlated four variables with extremes (2m temperature, total cloud cover, total precipitation, and volumetric soil moisture) and kept two variables uncorrelated (albedo and relative humidity). Examples of the synthetic data are shown in Figs. 6.4-6.9 and in Fig. 6.10.

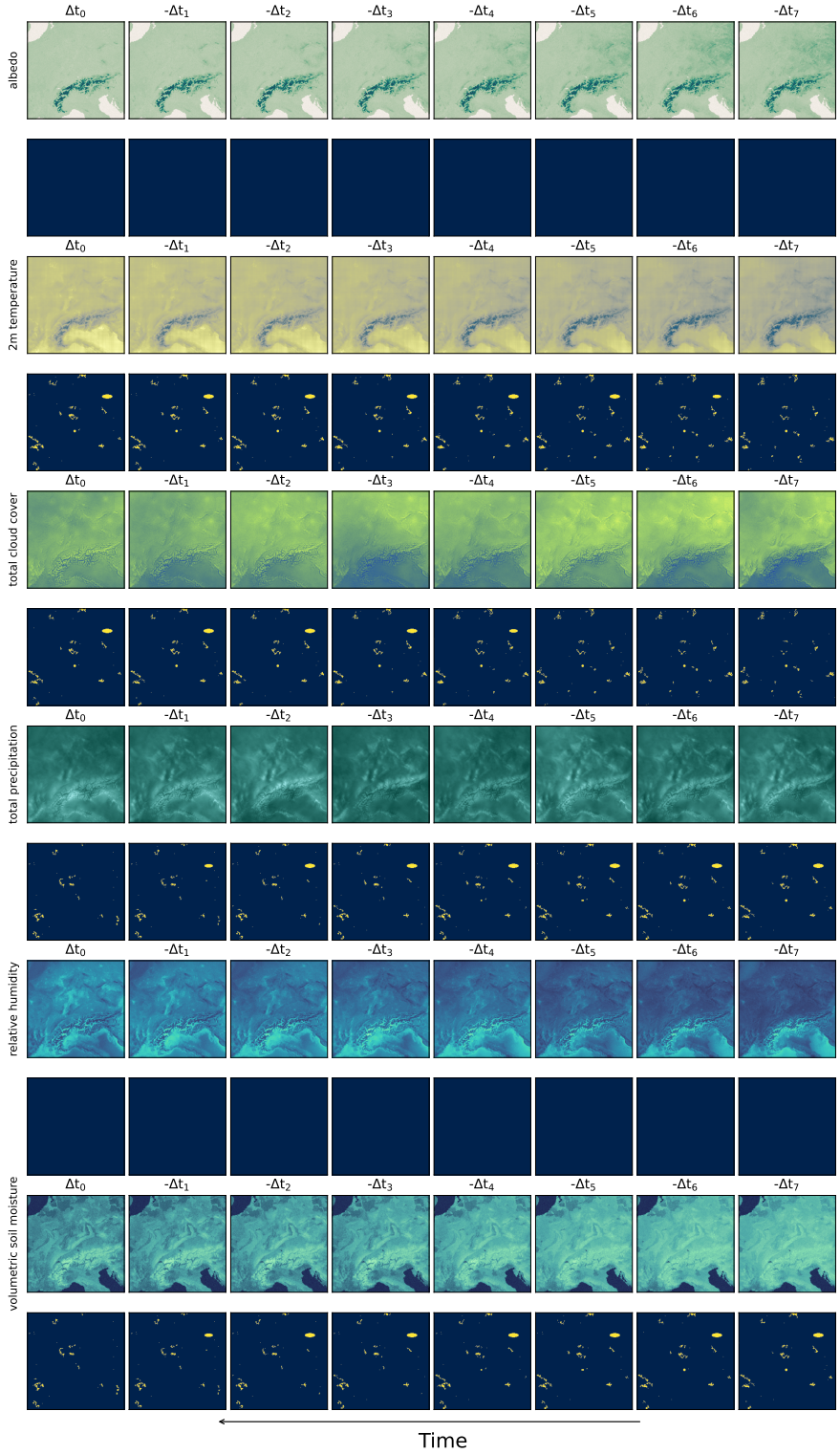


Figure 6.4: Examples of the synthetic CERRA reanalysis data described in Table 6.1. The drivers/anomalies ■ are visualized under each variable directly. Here, albedo and relative humidity are not correlated with the extremes.

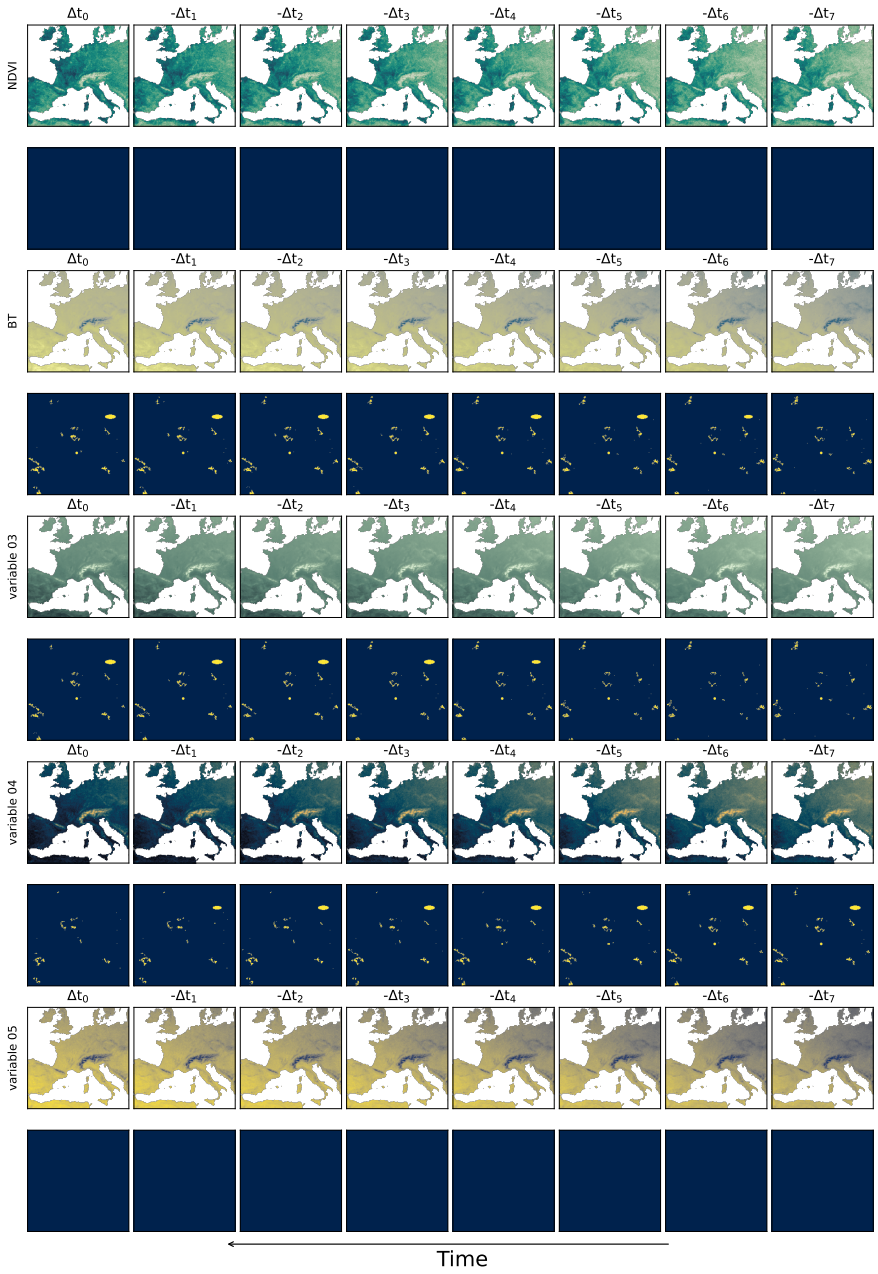


Figure 6.5: Examples of the synthetic NOAA remote sensing described in Table 6.2. The drivers/anomalies ■ are visualized under each variable directly. Here, NDVI and variables 05 are not correlated with the extremes.

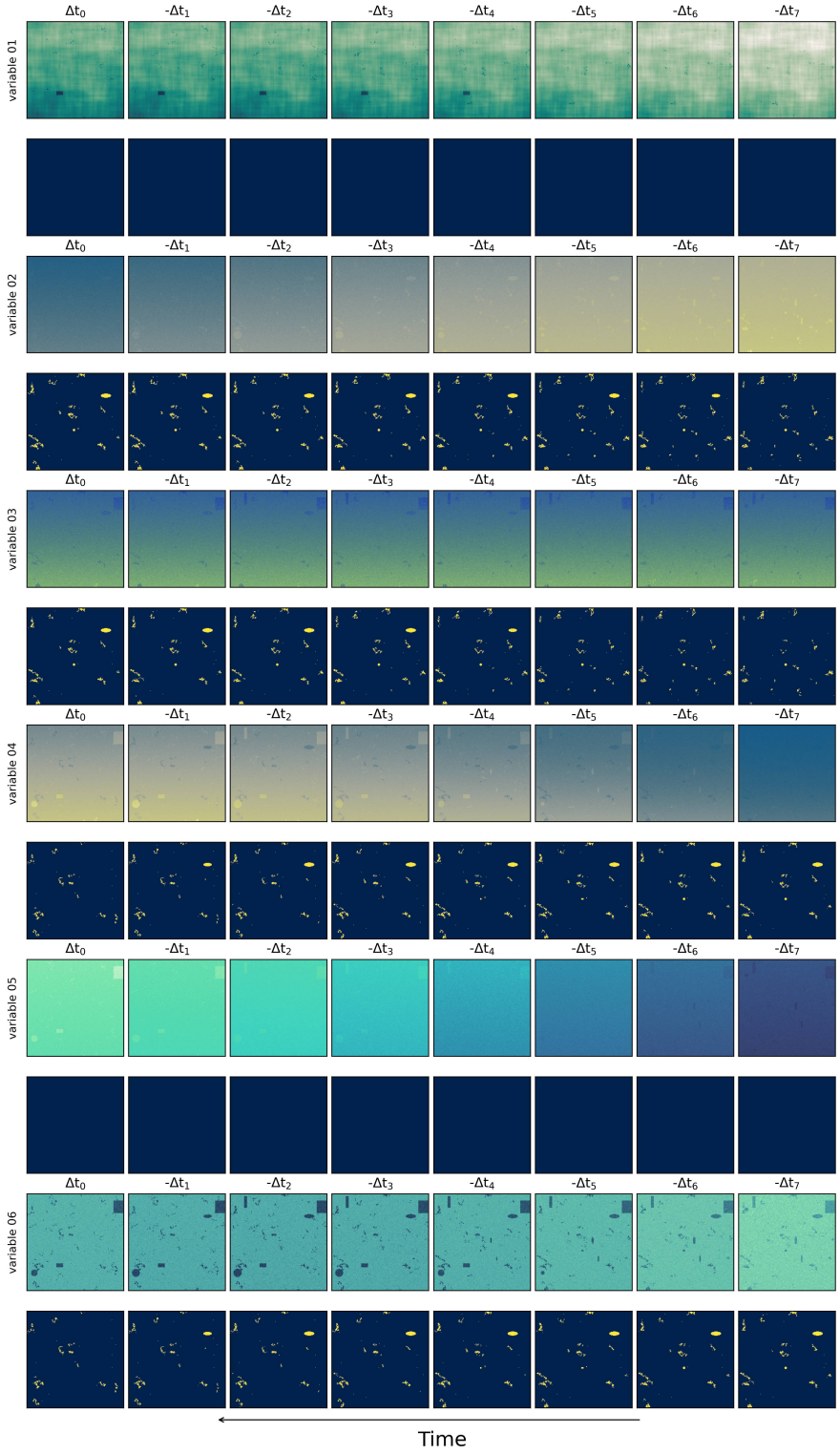


Figure 6.6: Examples of the synthetic artificial data described in Table 6.3. The drivers/anomalies ■ are visualized under each variable directly. Here, variables 01 and 05 are not correlated with the extremes.

Synthetic CERRA reanalysis								
Ind. variables	Dep. variables	Variables coupled with extreme	Dimension	Extreme events	$-\triangle T_a$	$+\Delta T_a$	% Extreme	% Correlated anomalies
6	0	4	{lon=200, lat=200, time=52×46}	{CubeEvent(n=200, sx=35, sy=35, sz=25), RandomWalkEvent(n=1100, s=125), LocalEvent(n=2600, sz=17), GaussianEvent(n=340, sx=35, sy=35, sz=25), OnsetEvent(n=25, sx=17, sy=17, os=0.98)}	9	4	1.16	1.69 % Random anomalous 1.32
Base	Dependency Weights		Noise	Random events	δ	kb	kn	ks
Albedo	-	-	WhiteNoise(meu=0, sigma=0.01)	{CubeEvent(n=320, sx=35, sy=35, sz=25), RandomWalkEvent(n=3000, s=125), LocalEvent(n=4000, sz=17), GaussianEvent(n=300, sx=35, sy=35, sz=25)}	-	0.30	0.20	0.50
2m Temperature	-	-	RedNoise(meu=0, sigma=0.90)	{OnsetEvent(n=18, sx=17, sy=17, os=0.98), RandomWalkEvent(n=1800, s=125), LocalEvent(n=160, sz=17), GaussianEvent(n=350, sx=35, sy=35, sz=25)}	+1	0.01	0.01	0.50
Total cloud cover	-	-	LaplaceNoise(meu=0, sigma=0.70, lambda=1)	{CubeEvent(n=300, sx=35, sy=35, sz=25), RandomWalkEvent(n=2000, s=125), LocalEvent(n=2800, sz=17), GaussianEvent(n=290, sx=35, sy=35, sz=25)}	-1	0.03	0.08	0.50
Total precipitation	-	-	WhiteNoise(meu=0, sigma=0.04)	{CubeEvent(n=320, sx=35, sy=35, sz=25), RandomWalkEvent(n=3000, s=125), LocalEvent(n=4000, sz=17), GaussianEvent(n=300, sx=35, sy=35, sz=25)}	-1	0.07	0.20	0.50
Relative humedity	-	-	CauchyNoise(meu=0, sigma=0.7)	{OnsetEvent(n=18, sx=17, sy=17, os=0.98), RandomWalkEvent(n=1800, s=125), LocalEvent(n=160, sz=17), GaussianEvent(n=350, sx=35, sy=35, sz=25)}	-	0.06	0.06	0.50
Volumetric soil moisture	-	-	WhiteNoise(meu=0., sigma=.017)	{CubeEvent(n=300, sx=35, sy=35, sz=25), RandomWalkEvent(n=2000, s=125), LocalEvent(n=2800, sz=17), GaussianEvent(n=290, sx=35, sy=35, sz=25)}	-1	0.10	0.10	0.50

Table 6.1: Configurations of the synthetic CERRA reanalysis.

Synthetic NOAA								
Ind. variables	Dep. variables	Variables coupled with extreme	Dimension	Extreme events	$-\triangle T_a$	$+\triangle T_a$	% Extreme	% Correlated anomalies
2	3	4	{lon=200, lat=200, time=52×46}	{CubeEvent(n=200, sx=35, sy=35, sz=25), RandomWalkEvent(n=1100, s=125), LocalEvent(n=2600, sz=17), GaussianEvent(n=340, sx=35, sy=35, sz=25), OnsetEvent(n=25, sx=17, sy=17, os=0.98)}	9	4	0.79	1.02 % Random anomalous 1.76
Base	Dependency Weights		Noise	Random events	δ	kb	kn	ks
NDVI	-	-	WhiteNoise(meu=0, sigma=0.30)	{CubeEvent(n=320, sx=35, sy=35, sz=25), RandomWalkEvent(n=3000, s=125), LocalEvent(n=4000, sz=17), GaussianEvent(n=300, sx=35, sy=35, sz=25)}	-	0.25	0.10	0.50
BT	-	-	WhiteNoise(meu=0, sigma=0.5)	{CubeEvent(n=300, sx=35, sy=35, sz=25), RandomWalkEvent(n=2000, s=125), LocalEvent(n=2800, sz=17), GaussianEvent(n=290, sx=35, sy=35, sz=25)}	+1	0.01	0.01	0.50
-	Quadratic Coupling()	Norm Weight()	WhiteNoise(meu=0, sigma=0.65)	-	-1	0.01	0.01	0.50
-	Linear Coupling()	Laplace Weight()	WhiteNoise(meu=0, sigma=.065)	-	-	0.01	0.01	0.50
-	Linear Coupling()	Laplace Weight()	WhiteNoise(meu=0., sigma=.065)	-	+1	0.01	0.01	0.50

Table 6.2: Configurations of the synthetic NOAA remote sensing.

Synthetic artificial								
Ind. variables	Dep. variables	Variables coupled with extreme	Dimension	Extreme events	$-\triangle T_a$	$+\triangle T_a$	% Extreme	% Correlated anomalies
3	3	4	{lon=200, lat=200, time=52×46}	{CubeEvent(n=200, sx=35, sy=35, sz=25), RandomWalkEvent(n=1100, s=125), LocalEvent(n=2600, sz=17), GaussianEvent(n=340, sx=35, sy=35, sz=25), OnsetEvent(n=25, sx=17, sy=17, os=0.98)}	9	4	1.24	1.81 % Random anomalous 2.93
Base	Dependency	Weights	Noise	Random events	δ	kb	kn	ks
SineBase(shift=0, amp=3, nOsc=46, latGrad=True)	-	-	RedNoise(meu=0, sigma=0.20)	{CubeEvent(n=320, sx=35, sy=35, sz=25), RandomWalkEvent(n=3000, s=125), LocalEvent(n=4000, sz=17), GaussianEvent(n=300, sx=35, sy=35, sz=25)}	-	0.35	0.35	0.35
CosineBase(shift=0, amp=3, nOsc=46, latGrad=True),	-	-	LaplaceNoise(meu=0, sigma=0.08, lambda=1)	{OnsetEvent(n=18, sx=17, sy=17, os=0.98), RandomWalkEvent(n=1800, s=125), LocalEvent(n=160, sz=17), GaussianEvent(n=350, sx=35, sy=35, sz=25)}	+1	0.35	0.35	0.35
ConstantBase(const=0, latGrad=True)	-	-	WhiteNoise(meu=0, sigma=0.07)	{CubeEvent(n=300, sx=35, sy=35, sz=25), RandomWalkEvent(n=2000, s=125), LocalEvent(n=2800, sz=17), GaussianEvent(n=290, sx=35, sy=35, sz=25)}	-1	0.90	0.90	0.90
-	Quadratic Coupling()	Norm Weight()	WhiteNoise(meu=0, sigma=0.65)	-	-1	0.35	0.35	0.35
-	Linear Coupling()	Laplace Weight()	WhiteNoise(meu=0, sigma=.065)	-	-	0.35	0.35	0.35
-	Linear Coupling()	Laplace Weight()	WhiteNoise(meu=0., sigma=.065)	-	-1	0.35	0.35	0.35

Table 6.3: Configurations of the synthetic artificial data.

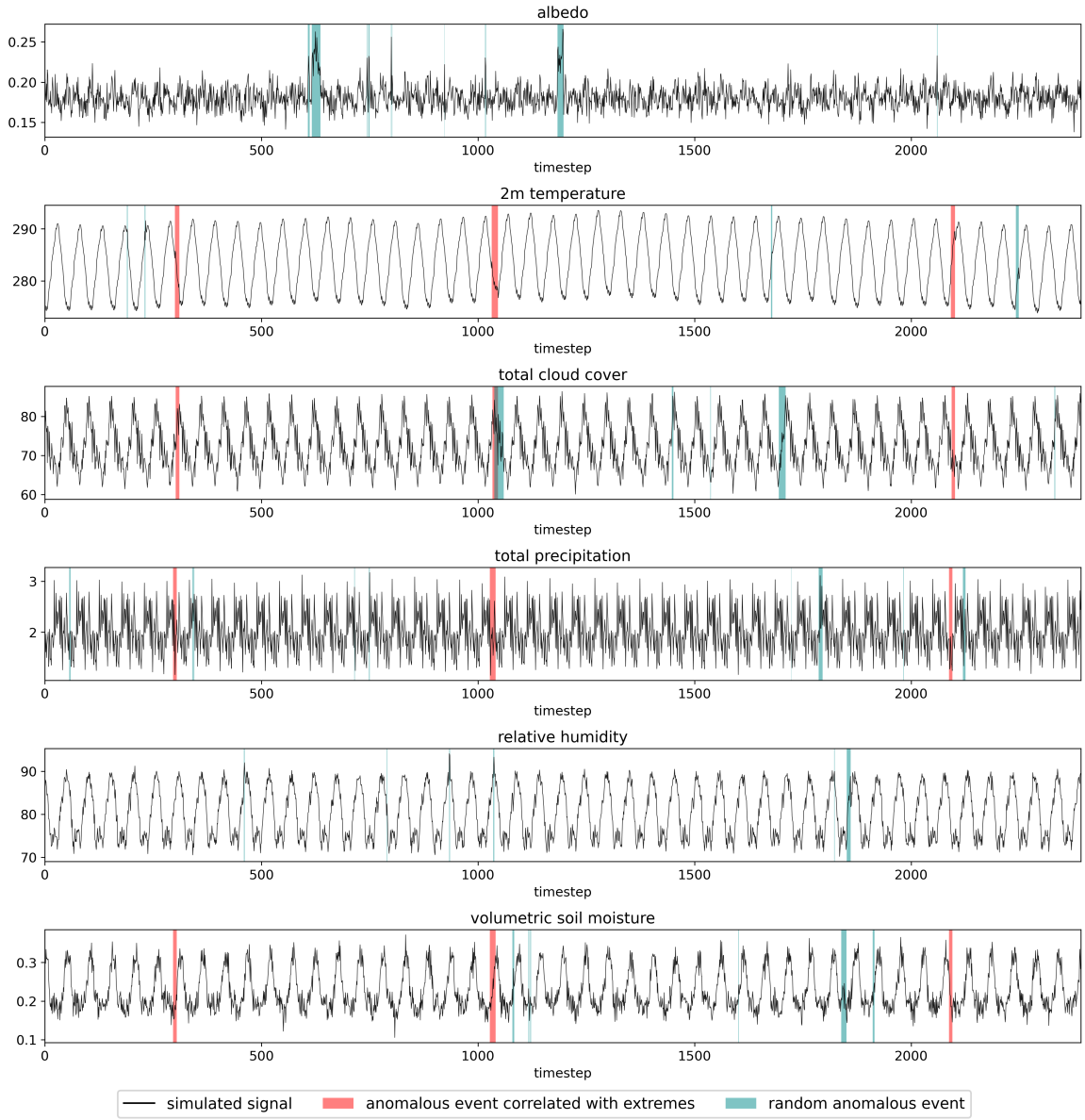


Figure 6.7: Visualization of the generated signals $\mathbf{\Phi}$ for 6 different variables from the synthetic CERRA reanalysis described in Table 6.1. The time series are shown for the location (lat $= 50$, lon $= 50$). ■ are the drivers/anomalies which are correlated with extremes, and ■ are random anomalies.

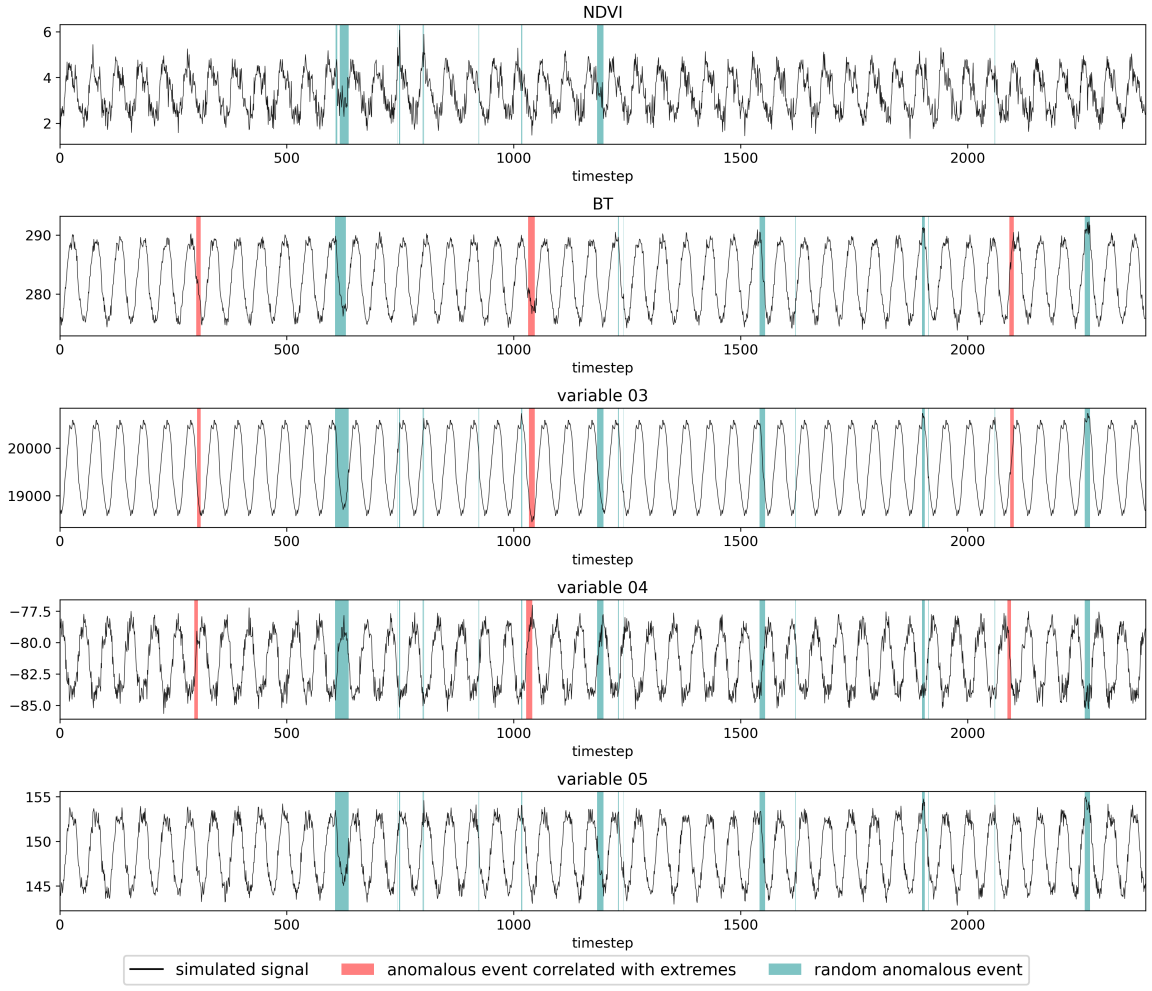


Figure 6.8: Visualization of the generated signals $\mathbf{\Phi}$ for 5 different variables from the synthetic NOAA data described in Table 6.2. The time series are shown for the location (lat $= 50$, lon $= 50$). ■ are the drivers/anomalies which are correlated with extremes, and ■ are random anomalies.

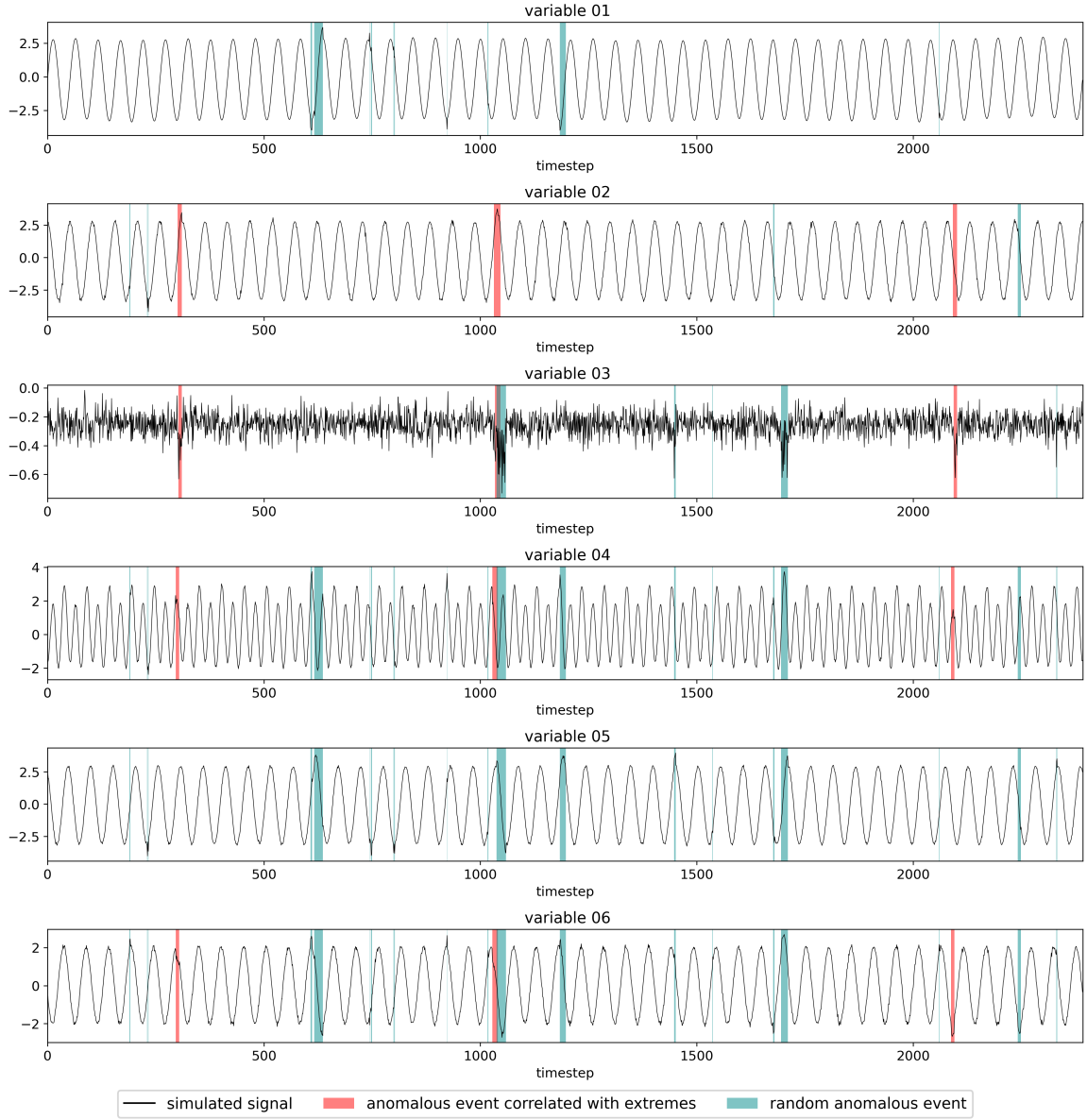


Figure 6.9: Visualization of the generated signals $\mathbf{\Phi}$ for 6 different variables from the synthetic artificial data described in Table 6.3. The time series are shown for the location (lat $= 50$, lon $= 50$). ■ are the drivers/anomalies which are correlated with extremes, and ■ are random anomalies.

		Validation		
	Algorithm	F1-score (↑)	IoU (↑)	OA (↑)
	Naive	0.4793	0.3152	0.9861
	Integrated Gradients I (*Sundararajan et al.*, 2017)	0.2415±0.0994	0.1412±0.0671	0.9218±0.0294
	Integrated Gradients II (*Sundararajan et al.*, 2017)	0.3123±0.0440	0.1858±0.0305	0.9526±0.0103
One-Class	OCSVM (*Schölkopf et al.*, 1999)	0.2821±0.0249	0.1644±0.0167	0.9564±0.0016
One-Class	IF (*Liu et al.*, 2012)	0.3499±0.0056	0.2128±0.0042	0.9716±0.0002
One-Class	SimpleNet (*Liu et al.*, 2023b)	0.7531±0.0007	0.6039±0.0010	0.9920±0.0001
Rec.	STEALNet (*Astrid et al.*, 2021)	0.5598±0.0090	0.3887±0.0086	0.9847±0.0003
Rec.	UniAD (*You et al.*, 2022)	0.4753±0.0017	0.3118±0.0014	0.9744±0.0002
MIL	DeepMIL (*Sultani et al.*, 2018)	0.7068±0.0161	0.5468±0.0191	0.9922±0.0003
MIL	ARNet (*Wan et al.*, 2020)	0.7292±0.0085	0.5739±0.0106	0.9926±0.0001
MIL	RTFM (*Tian et al.*, 2021)	0.6009±0.0031	0.4295±0.0031	0.9834±0.0003
	Ours*	0.8278±0.0053	0.7063±0.0078	0.9951±0.0002
	Ours†	**0.8345**±0.0037	**0.7160**±0.0054	**0.9952**±0.0001
		Testing		
	Algorithm	F1-score (↑)	IoU (↑)	OA (↑)
	Naive	0.5124	0.3445	0.9837
	Integrated Gradients I (*Sundararajan et al.*, 2017)	0.2314±0.0705	0.1327±0.0465	0.9158±0.0225
	Integrated Gradients II (*Sundararajan et al.*, 2017)	0.3034±0.0427	0.1795±0.0294	0.9419±0.0122
One-Class	OCSVM (*Schölkopf et al.*, 1999)	0.2998±0.0226	0.1766±0.0154	0.9491±0.0019
One-Class	IF (*Liu et al.*, 2012)	0.3716±0.0067	0.2284±0.0051	0.9661±0.0003
One-Class	SimpleNet (*Liu et al.*, 2023b)	0.7350±0.0024	0.5811±0.0030	0.9891±0.0002
Rec.	STEALNet (*Astrid et al.*, 2021)	0.5774±0.0095	0.4060±0.0093	0.9822±0.0003
Rec.	UniAD (*You et al.*, 2022)	0.4923±0.0041	0.3265±0.0036	0.9718±0.0005
MIL	DeepMIL (*Sultani et al.*, 2018)	0.7154±0.0160	0.5572±0.0192	0.9909±0.0004
MIL	ARNet (*Wan et al.*, 2020)	0.7368±0.0086	0.5834±0.0108	0.9913±0.0002
MIL	RTFM (*Tian et al.*, 2021)	0.6188±0.0028	0.4481±0.0030	0.9812±0.0002
	Ours*	0.8044±0.0070	0.6728±0.0097	**0.9929**±0.0004
	Ours†	**0.8065**±0.0024	**0.6758**±0.0033	**0.9929**±0.0001

*Video Swin Transformer backbone (*Liu et al.*, 2022b) †Mamba backbone (*Gu and Dao*, 2024)

Table 6.4: Driver detection results on the synthetic CERRA reanalysis. The best performance on each metric is highlighted in a bold text. (±) denotes the standard deviation for 3 runs.

6.4.3 Experiments on Synthetic Data

The results on the synthetic CERRA reanalysis are reported in Table 6.4 and the results on synthetic NOAA and the artificial data are shown in Tables 6.5 and 6.6, respectively.

6.4.3.1 Quantitative Results

For synthetic CERRA, the quantitative results are shown in Table 6.4. The naive baseline is impacted by two main issues; first by the time delay between drivers and extreme events, and second not all variables are correlated with the extremes. The second issue affects the one-class and reconstruction-based baselines where they suffer mostly from false positives. In fact, both integrated gradients models achieve high F1-scores for detecting extremes (93.32 for Integrated Gradients I and 93.80 for Integrated Gradients II), but they have worse performance on identifying the drivers. SimpleNet is trained with our model as a feature extractor which explains its good performance. However, SimpleNet showed a drop of performance when it is tested on other datasets (Tables 6.5 and 6.6). Among the reconstruction-based approach, STEALNet outperforms UniAD. This is probability because STEALNet exploits more weakly supervision information during training by maximizing the reconstruction loss for locations with extreme flags. MIL-based baselines are more suitable for the task. Finally, our model consistently outperforms the baselines on all metrics.

In Tables 6.5 and 6.6, we report the results on the synthetic NOAA remote sensing and artificial data. We noticed a drop of performance for all models for the synthetic artificial data (Table 6.6). This explained as the ratio of anomalies is higher than the other two datasets. Second, we used for this artificial dataset red noise and quadratic coupling to generate dependent base signals. This makes this dataset harder for training. SimpleNet exhibits a dramatic dropped out of performance when it is tested on these two synthetic datasets. This illustrates that the model performance is highly dependent on the dataset and the backbone the model was trained on. Our model still outperforms all baselines on these datasets. Note that for the Mamba backbone used here, we do not use bidirectional SSM. An improvement is expected with bidirectional Mamba, as we will see in Chapter 7.

6.4.3.2 Qualitative Results

Qualitative samples in comparison with baselines are presented in Fig. 6.10. The qualitative examples indicate that our model and the MIL-based baselines except RTFM are capable of learning which variables are correlated with the extremes. The main weakness of RTFM is the reliance on feature magnitudes and the cross attention module (see Section 6.4.1.3 and Table 6.9), which make it more prone to produce false positives. Other baselines predict incorrect relations between the variables and extremes. Regarding the explainable AI methods, when we add more interactions between the variables (Integrated Gradients II), the gradients tend to omit some variables (soil moisture). Both integrated gradients models have also difficulties with the synthetic t2m, which includes red noise by design. These results demonstrate that networks that predict the extremes directly from the input variables utilize much more information even when it is not correlated with an extreme. It is thus beneficial to introduce a bottleneck into the network that enforces the network to explicitly identify drivers of extremes.

		Validation		
	Algorithm	F1-score (↑)	IoU (↑)	OA (↑)
	Naive	0.4747	0.3112	0.9917
One-Class	OCSVM (*Schölkopf et al.*, 1999)	0.3794±0.1311	0.2421±0.0993	0.9860±0.0030
One-Class	IF (*Liu et al.*, 2012)	0.4414±0.0118	0.2834±0.0097	0.9868±0.0001
One-Class	SimpleNet (*Liu et al.*, 2023b)	0.5694±0.0020	0.3980±0.0019	0.9929±0.0002
Rec.	STEALNet (*Astrid et al.*, 2021)	0.5697±0.0086	0.3984±0.0084	0.9906±0.0001
Rec.	UniAD (*You et al.*, 2022)	0.4524±0.0358	0.2930±0.0294	0.9865±0.0021
MIL	DeepMIL (*Sultani et al.*, 2018)	0.7177±0.0038	0.5597±0.0046	0.9955±0.0001
MIL	ARNet (*Wan et al.*, 2020)	0.7106±0.0046	0.5511±0.0056	0.9953±0.0001
MIL	RTFM (*Tian et al.*, 2021)	0.6030±0.0026	0.4316±0.0027	0.9892±0.0002
	Ours*	**0.8193**±0.0023	**0.6940**±0.0036	**0.9969**±0.0001
	Ours†	0.8144±0.0047	0.6870±0.0066	**0.9969**±0.0001
		Testing		
	Algorithm	F1-score (↑)	IoU (↑)	OA (↑)
	Naive	0.5107	0.3429	0.9895
One-Class	OCSVM (*Schölkopf et al.*, 1999)	0.3925±0.1257	0.2517±0.0967	0.9826±0.0037
One-Class	IF (*Liu et al.*, 2012)	0.4513±0.0130	0.2915±0.0109	0.9849±0.0001
One-Class	SimpleNet (*Liu et al.*, 2023b)	0.5724±0.0031	0.4010±0.0030	0.9908±0.0002
Rec.	STEALNet (*Astrid et al.*, 2021)	0.5865±0.0088	0.4150±0.0088	0.9883±0.0002
Rec.	UniAD (*You et al.*, 2022)	0.4699±0.0421	0.3081±0.0354	0.9836±0.0029
MIL	DeepMIL (*Sultani et al.*, 2018)	0.7202±0.0031	0.5628±0.0038	0.9942±0.0001
MIL	ARNet (*Wan et al.*, 2020)	0.7143±0.0051	0.5556±0.0062	0.9940±0.0001
MIL	RTFM (*Tian et al.*, 2021)	0.6191±0.0021	0.4483±0.0023	0.9870±0.0002
	Ours*	**0.8255**±0.0025	**0.7028**±0.0036	**0.9961**±0.0001
	Ours†	0.8207±0.0045	0.6959±0.0065	0.9960±0.0001

*Video Swin Transformer backbone (*Liu et al.*, 2022b) †Mamba backbone (*Gu and Dao*, 2024)

Table 6.5: Driver detection results on the synthetic NOAA remote sensing. The best performance on each metric is highlighted in a bold text. (±) denotes the standard deviation for 3 runs.

		Validation		
	Algorithm	F1-score (↑)	IoU (↑)	OA (↑)
	Naive	0.4689	0.3063	0.9849
One-Class	OCSVM (*Schölkopf et al.*, 1999)	0.2796±0.0176	0.1627±0.0120	0.9777±0.0011
One-Class	IF (*Liu et al.*, 2012)	0.2855±0.0121	0.1666±0.0082	0.9763±0.0002
One-Class	SimpleNet (*Liu et al.*, 2023b)	0.3457±0.0054	0.2090±0.0039	0.9756±0.0004
Rec.	STEALNet (*Astrid et al.*, 2021)	0.5640±0.0078	0.3928±0.0076	0.9833±0.0003
Rec.	UniAD (*You et al.*, 2022)	0.4948±0.0160	0.3288±0.0141	0.9766±0.0011
MIL	DeepMIL (*Sultani et al.*, 2018)	0.2018±0.2367	0.1338±0.1645	0.7133±0.2053
MIL	ARNet (*Wan et al.*, 2020)	0.4898±0.0530	0.3259±0.0455	0.9882±0.0006
MIL	RTFM (*Tian et al.*, 2021)	0.5990±0.0031	0.4275±0.0032	0.9827±0.0004
	Ours*	**0.7020**±0.0043	**0.5408**±0.0051	**0.9890**±0.0003
	Ours†	0.6663±0.0541	0.5019±0.0591	0.9881±0.0012
		Testing		
	Algorithm	F1-score (↑)	IoU (↑)	OA (↑)
	Naive	0.5109	0.3431	0.9828
One-Class	OCSVM (*Schölkopf et al.*, 1999)	0.3305±0.0129	0.1980±0.0093	0.9742±0.0012
One-Class	IF (*Liu et al.*, 2012)	0.3415±0.0125	0.2060±0.0091	0.9719±0.0003
One-Class	SimpleNet (*Liu et al.*, 2023b)	0.4118±0.0035	0.2593±0.0028	0.9794±0.0009
Rec.	STEALNet (*Astrid et al.*, 2021)	0.5823±0.0092	0.4109±0.0092	0.9811±0.0004
Rec.	UniAD (*You et al.*, 2022)	0.5249±0.0125	0.3560±0.0116	0.9757±0.0009
MIL	DeepMIL (*Sultani et al.*, 2018)	0.1875±0.2134	0.1204±0.1439	0.7108±0.2048
MIL	ARNet (*Wan et al.*, 2020)	0.4475±0.0553	0.2898±0.0449	0.9853±0.0007
MIL	RTFM (*Tian et al.*, 2021)	0.6141±0.0046	0.4431±0.0048	0.9807±0.0004
	Ours*	**0.7033**±0.0071	**0.5424**±0.0084	**0.9874**±0.0008
	Ours†	0.6764±0.0595	0.5139±0.0660	0.9871±0.0018

*Video Swin Transformer backbone (*Liu et al.*, 2022b) †Mamba backbone (*Gu and Dao*, 2024)

Table 6.6: Driver detection results on the synthetic artificial data. The best performance on each metric is highlighted in a bold text. (±) denotes the standard deviation for 3 runs.

Figure 6.10: Qualitative results on the synthetic CERRA reanalysis from the test set at time step 2160. ■ is the prediction, ■ is the ground truth, and ■ is the false positive. Albedo and relative humidity are not correlated with extremes, meaning that they do not contain drivers, but only random anomalies.

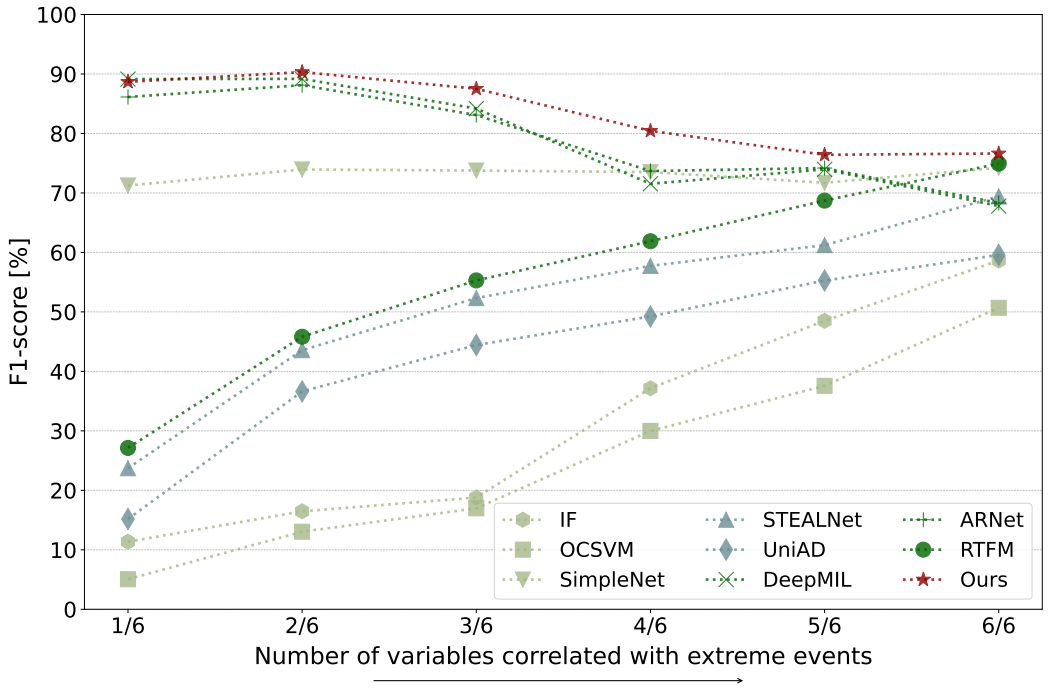


Figure 6.11: F1-score with different correlation settings between the input variables and extreme events.

6.4.3.3 Performance on Easy-to-Hard Correlation Settings

We conduct an additional experiment to assess the model performance in relation to the correlation setup between the variables and extremes. We generate a synthetic CERRA dataset starting with only one correlated variable with the target extreme. We then generate different versions of the dataset by increasing the number of correlated variables with the extremes up to 100%. This analysis allows us to point out the strengths and weaknesses of the comparative models for different scenarios and where our model becomes more effective, as well as where it could struggle most. The results are shown in Fig. 6.11. One-class, reconstruction-based and RTFM baselines benefit with increasing the number of correlated variables. In case of 6/6, the task reduces to an anomaly detection task. The performance of our model and MIL-based baselines generally decreases when the number of correlated variables increases as the task of finding all correlated anomalies becomes harder. Nevertheless, our approach performs best in all settings.

6.4.3.4 Computational Efficiency

In Table 6.7, we report the inference time with a fixed input of 6 variables, 8 days and 200×200 spatial resolution. STEALNet is based on a 3D CNN auto-encoder without a self-attentions which explains its efficiency but on the cost of accuracy. In the last row, we show the estimated time when we discard the classification head to detect extreme events and only use the model for anomaly detection.

Algorithm	GPU[1] (sec)	Parameters (M)
SimpleNet	0.156 $\pm$0.003	0.203
STEALNet	0.003 $\pm$0.000	6.005
UniAD	3.733 $\pm$0.012	3.674
DeepMIL	0.193 $\pm$0.008	0.285
ARNet	0.191 $\pm$0.003	0.285
RTFM	0.257 $\pm$0.001	0.319
Ours	0.132 $\pm$0.000	0.479
Ours[2]	0.122 $\pm$0.000	0.145

[1]NVIDIA GeForce RTX 3090 GPU

[2]Our model is used only for driver/anomaly detection

Table 6.7: Inference time in seconds for our model and other DL baselines.

The training on the real-world data for EUR-11 took about $\sim$ 21 hours with a Swin Transformer model, $K = 16$, and 4 NVIDIA A100 GPUs. Table 6.8 gives a rough estimation for training on the synthetic CERRA for 1 epoch. SimpleNet was trained with a pretrained backbone. The training time includes some postprocessing to compute metrics on the training set. The time might also differ depending on the I/O during training and the number of available workers.

6.4.4 Ablation Studies

In this section, we do ablation analyses on different aspects of the proposed model. All experiments are done on the validation set of synthetic CERRA reanalysis, unless otherwise stated and evaluated with the F1-scores for drivers anomalies and extreme events detection.

Algorithm	Time (min)	GPU
SimpleNet	~ 2	A100
STEALNet	~ 1	A100
UniAD	~ 11	$4 \times$ A100
DeepMIL	~ 13	A40
ARNet	~ 13	A40
RTFM	~ 20	A40
Ours	~ 8	A40

Table 6.8: Training time on the synthetic CERRA for 1 epoch.

6.4.4.1 Feature Extractor (f_θ)

In Table 6.9, we show the benefit of having independent feature extractors for driver detection. In a first experiment, we share the feature extractor f_θ among the variables. The performance is worst. Second, we replaced the temporal attention by a cross attention between the variables similar to *Nguyen et al.* (2023) and *Lessig et al.* (2023) where each variable performs a cross attention with the other variables. We can see a drop of performance for the second experiment. We noticed that anomalies propagate between variables when adding connections in the feature extraction stage. This also explains the poor performance of RTFM compared to other MIL baselines. The best performance is shown for the proposed setup (last row in Table 6.9), which also shows the benefit for the temporal attention.

Temporal-attention	Cross-attention	Shared-(f_θ)	F1-score (↑)
✓	✗	✓	0.6959 | 0.9042
✗	✓	✗	0.6718 | 0.9378
✗	✗	✗	0.8239 | 0.9197
✓	✗	✗	**0.8278 | 0.9245**

Table 6.9: Ablation studies on the key model architecture from the validation set. The metric is F1 on the driver | extreme drought detection. The best performance on each metric is highlighted in a bold text.

In Table 6.10, we evaluate three types of feature extractors, 3D CNN, Video Swin Transformer (*Liu et al.*, 2022b), and SSM Vision Mamba (*Gu and Dao*, 2024) with local scans. All models are trained from scratch. To build the 3D CNN model, we replaces the attention blocks in Swin Transformer with 3D convolutions and keep the overall architecture. For Mamba, we use local scans. Vision Mamba backbones replace the attention module with a linear selective state space model which provides an efficient alternative for Video Swin Transformer (Section 2.1.7). From Table 6.10 we notice that Swin Transformer achieves better results compared to 3D CNN. The parameters are also less compared to 3D CNN. The results also indicate that using Mamba instead of Swin Transformer achieves similar or better results when less parameters are used. In contrast to Swin Transformer, Mamba can be scaled to the global scale. However, we have used Swin Transformer in all experiments to have a fair comparison with the baselines due to its commonality as a backbone.

Table 6.10 shows that increasing the model parameters still does not show a sign of overfitting.

Backbone f_θ	Hidden dimension (K)	Parameters	F1-score (↑)
3D CNN	8	63k	0.5715 | 0.9121
Video Swin Transformer	8	19k	0.8122 | 0.9116
Mamba	8	15k	0.8215 | 0.9018
3D CNN	16	250k	0.7093 | 0.9375
Video Swin Transformer	16	62k	0.8278 | 0.9245
Mamba	16	56k	0.8345 | 0.9312
3D CNN	32	998k	0.8495 | 0.9343
Video Swin Transformer	32	230k	0.8414 | 0.9312
Mamba	32	214k	0.8400 | 0.9343

Table 6.10: Ablation study on the backbone f_θ used for feature extraction. The metric is F1-score on the driver | extreme drought detection.

6.4.4.2 Temporal Resolution T

Table 6.11 evaluates the impact of the temporal resolution on driver and extreme detection. T=6 provides a good balance between driver and extreme detection.

T	1	4	6	8
F1-score (↑)	0.8338 | 0.6925	0.8136 | 0.9068	**0.8278 | 0.9245**	0.7733 | 0.9030

Table 6.11: Ablation studies on the temporal resolution from the validation set. The metric is F1 on the driver | extreme drought detection. The best performance on each metric is highlighted in a bold text.

6.4.4.3 Loss Functions

As shown in Table 6.12, $\mathcal{L}_{(quantize)}$ and $\mathcal{L}_{(driver)}$ are essential for training. As other quantization models, ours can not be trained without $\mathcal{L}_{(quanitze)}$, which ensures that the outputs of f_θ do not grow and commit to the binary embedding. $\mathcal{L}_{(driver)}$ unifies the representation of drivers for all variable as $q = 1$, which boosts the extreme detection. Moreover, the results demonstrate the impact of using V+1 3D CNNs (multi-head) instead of one for $\mathcal{L}_{(extreme)}$. If a single 3D CNN is used, drivers are only identified in a small subset of variables.

$\mathcal{L}_{(quantize)}$	$\mathcal{L}_{(driver)}$	$\mathcal{L}_{(extreme)}$ multi-head	F1-score (↑)
✓	✗	✗	0.2999 | 0.4894
✓	✗	✓	0.0251 | 0.9288
✓	✓	✗	0.3123 | 0.6640
✓	✓	✓	**0.8278 | 0.9245**

Table 6.12: Ablation studies on the loss function from the validation set. The metric is F1 on the driver | extreme drought detection. The best performance on each metric is highlighted in a bold text.

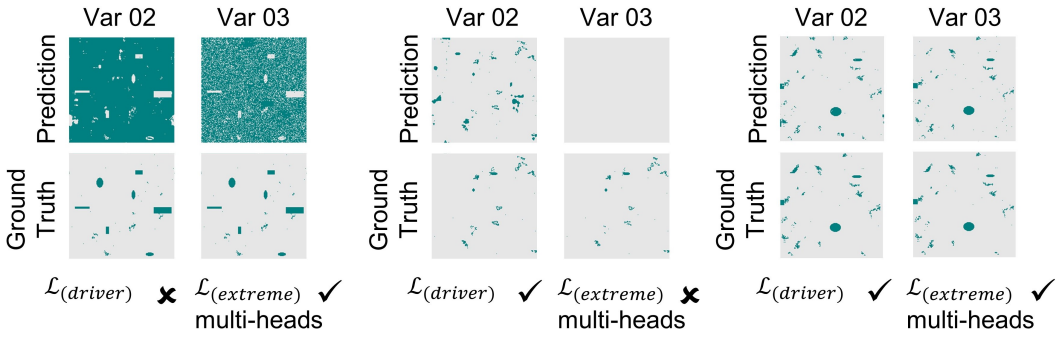


Figure 6.12: Supplementary to the ablation study in Table 6.12.

Without the loss $\mathcal{L}_{(driver)}$, the detection of drivers/anomalies is not reliable since pixels at regions and intervals where no extreme event occurred can be assigned to $z_{q=1}$ (driver) as well. In case of $\mathcal{L}_{(extreme)}$ multi-heads, we observe that anomalies are identified in a small subset of variables because the network omits some variables if there is a correlation with other variables. Please see Fig. 6.12. In case of a single head such flips occur less often, but they can occur. If the loss $\mathcal{L}_{(driver)}$ is used, such flips cannot occur and the multi-head improves both F1-scores by a large margin. When comparing rows 2 and 4 in Table 6.12, there is a slight decrease in extreme prediction but a large improvement in anomaly detection. Note that there is always a trade-off between extreme and anomaly detection. The anomalies generate a bottleneck of information. When more information goes through the bottleneck the better the extreme prediction gets. Without any anomaly detection, the extreme prediction is best as shown in Table 6.18, but the increase in F1 score is only moderate. This is also visible in rows 2 and 3 in Table 6.9. Cross-attention improves extreme prediction, but it hurts the detection of anomalies since the information is propagated between the variables.

Loss function for quantization. This part studies the loss function $\mathcal{L}_{(quantize)}$ used in Eq. (6.5). We denote the term $\|\mathbf{Z}_l - \text{sg}(\text{sign}(\mathbf{Z}_l))\|_2^2$ as $\mathcal{L}_{(commit)}$, $H[\mathbb{E}(\text{sign}(\mathbf{Z}_l))]$ as $\mathcal{L}_{(div)}$, and $\mathbb{E}[H(\text{sign}(\mathbf{Z}_l))]$ as $\mathcal{L}_{(ent)}$. As seen from Table 6.13, $\mathcal{L}_{(commit)}$ is essential for training to prevent the input for the quantization layer from growing. While $\mathcal{L}_{(ent)}$ and $\mathcal{L}_{(div)}$ (rows 4, 5, and 7) improve the results compared to the model with $\mathcal{L}_{(commit)}$ (second row).

$\mathcal{L}_{(ent)}$	$\mathcal{L}_{(commit)}$	$\mathcal{L}_{(div)}$	F1-score (↑)
✓	✗	✗	0.0000 | 0.0000
✗	✓	✗	0.8186 | 0.9118
✗	✗	✓	0.0000 | 0.0000
✓	✓	✗	0.8187 | 0.9148
✗	✓	✓	0.8226 | 0.9230
✓	✗	✓	0.0000 | 0.0000
✓	✓	✓	**0.8278 | 0.9245**

Table 6.13: Ablation studies on the loss function $\mathcal{L}_{(quantize)}$ in Eq. (6.5). The metric is F1-score on the driver | extreme drought detection. The best performance on each metric is highlighted in a bold text.

Loss function for extreme events prediction. Due to class imbalanced, we add class weighting applied to the binary cross entropy loss $\mathcal{L}_{(extreme)}$ to predict extremes. The weighting is based on the logarithm of the inverse square roots of the relative class frequencies in the batch. The weighted loss function achieves better results as shown in Table 6.14.

$\mathcal{L}_{(extreme)}$	F1-score (↑)
unweighted $\mathcal{L}_{(extreme)}$	0.8114 | 0.9091
weighted $\mathcal{L}_{(extreme)}$	**0.8278 | 0.9245**

Table 6.14: Ablation studies on the weighted loss function $\mathcal{L}_{(extreme)}$ in Eq. (6.4). The metric is F1-score on the driver | extreme drought detection. The best performance on each metric is highlighted in a bold text.

Weighting parameters in the main loss function. We study the role of the weighting parameters in the main loss function in Eq. (6.7). We can observe from Table 6.15 that $\lambda_{(driver)}$ plays a crucial role in anomaly detection. Having $\lambda_{(driver)}$ small makes the model less constrained and reduces supervision on where extremes were reported in the training data. Bigger values of $\lambda_{(driver)}$ make the model more constrained to identify drivers near or in overlap with the extremes and thus reduce the overall performance. We can also observe that the model is less sensitive to $\lambda_{(ent)}$ compared to other parameters. We selected the final default weighting parameters in the experiments based on the average performance on both driver and extreme events prediction.

(a)

$\lambda_{(driver)}$	F1-score (↑)
1	0.0245 | 0.5807
10	0.0334 | 0.7573
100	**0.8278 | 0.9245**
200	0.8085 | 0.9169
1000	0.7994 | 0.8644

(b)

$\lambda_{(commit)}$	F1-score (↑)
1.0	0.7981 | 0.9202
1.5	0.8143 | 0.9196
2.0	0.8227 | 0.9211
3.0	**0.8278 | 0.9245**
4.0	0.8276 | 0.9189

(c)

$\lambda_{(ent)}$	F1-score (↑)
0.01	0.8252 | 0.9261
0.1	**0.8278 | 0.9245**
1.0	0.8272 | 0.9246

(d)

$\lambda_{(div)}$	F1-score (↑)
0.01	0.7751 | 0.9171
0.1	**0.8278 | 0.9245**
0.5	0.8248 | 0.9221

Table 6.15: Sensitivity studies on the weighting parameters used in the main loss function in Eq. (6.7). The metric is mean F1-score on the driver | extreme drought detection for 3 random seeds. The best performance on each metric is highlighted in a bold text.

6.4.4.4 Quantization Layer

In Table 6.16, we study the performance of our model with different vector quantization algorithms. The first row *Threshold (Tanh)* represents a simple straight through estimator. For this quantization, we first map the input into a scalar value followed by a Tanh activation. Then we set positive values to be anomalies ($q = 1$). In *Random Quantization (RQ)* (*Chiu et al.*, 2022) the input is projected with a randomly initialized weights and then compared with a randomly initialized codes. *Vector Quantization (VQ)* is the standard quantizer which uses an Euclidean distance (*Van Den Oord et al.*, 2017) or a cosine similarity (*Yu et al.*, 2022a). We further add an orthogonality loss for VQ similar

to (*Shin et al.*, 2023). *Finite Scalar Quantization (FSQ)* maps the input into a bounded scalar value. The code is then assigned based on the rounded value in the discrete space. As seen, *Lockup free quantization (LFQ)* (*Yu et al.*, 2024) performs the best. We speculate that LFQ does not need to learn the code vectors which simplifies the task of quantization.

Quantization Layer	F1-score (↑)
Threshold (Tanh)	0.7007 | 0.8498
RQ (Euclidean distance) (*Chiu et al.*, 2022)	0.7152 | 0.8723
VQ (Cosine similarity) (*Yu et al.*, 2022a)	0.7883 | 0.8898
VQ (Cosine similarity + Orthogonality) (*Yu et al.*, 2022a; *Shin et al.*, 2023)	0.7854 | 0.8683
VQ (Euclidean distance) (*Van Den Oord et al.*, 2017)	0.7742 | 0.8727
VQ (Euclidean distance + Orthogonality) (*Van Den Oord et al.*, 2017; *Shin et al.*, 2023)	0.7926 | 0.8828
FSQ (*Mentzer et al.*, 2024)	0.7657 | 0.8865
LFQ (*Yu et al.*, 2024)	**0.8278 | 0.9245**

Table 6.16: Ablation studies on the quantization layer. The metric is F1-score on the driver | extreme drought detection. The best performance on each metric is highlighted in a bold text.

6.4.5 Experiments on Real-world Data

We evaluate our model on two reanalysis data with diverse geographical and climate regions. The input for the experiment is the normalized mean and standard deviation of each week. We exclude pixels over water surfaces, desert, and snow. In addition to the quantitative evaluation on the synthetic data, we aim to verify our method considering the following aspects:

6.4.5.1 Quantitative Results

We expect that the developed model can identify drivers in real-world scenarios. We demonstrate this by measuring how well the model can predict extreme agricultural droughts from the identified drivers. To verify this, we report in Table 6.17 the prediction accuracy on extreme agricultural droughts. The results show that using the identified drivers as input, the model can predict extreme agricultural droughts across different regions and datasets. Note that compared to the synthetic dataset, the real-world drought prediction is much more difficult.

The performance depends on the type and ratio of extremes, spatio-temporal resolution, the quality and consistency between the remote sensing and the reanalysis data. F1 scores substantially increase when the threshold on VHI is increased (see Table 6.20 in Section 6.4.5.5). Note that it is not required to predict all extremes in order to learn specific relations from the predicted events.

6.4.5.2 Extreme Detection Without Anomaly Detection

We trained the model without the quantization step, meaning without driver detection. In other words, we remove the driver detection step and trained the model to predict extreme droughts directly from the extracted features with one classification head. This can be considered as an upper bound on the extreme detection accuracy since there is no information reduction by the quantization. Table 6.18 shows the improvement in performance on both the synthetic and real-world datasets. Adding

		Validation		
Dataset	Region	F1-score (↑)	IoU (↑)	OA (↑)
CERRA	Europe	0.2231±0.0074	0.1256±0.0047	0.9045±0.0105
ERA5-Land	Europe	0.3187±0.0039	0.1896±0.0028	0.9545±0.0025
	Africa	0.2249±0.0042	0.1267±0.0027	0.8559±0.0148
	North America	0.2739±0.0063	0.1587±0.0042	0.9384±0.0013
	South America	0.2930±0.0152	0.1717±0.0104	0.8931±0.0033
	Central Asia	0.2099±0.0043	0.1173±0.0027	0.9538±0.0004
	East Asia	0.1882±0.0072	0.1039±0.0044	0.9310±0.0042
		Testing		
Dataset	Region	F1-score (↑)	IoU (↑)	OA (↑)
CERRA	Europe	0.2863±0.0145	0.1671±0.0098	0.8913±0.0144
ERA5-Land	Europe	0.2152±0.0086	0.1206±0.0054	0.9584±0.0021
	Africa	0.1853±0.0048	0.1021±0.0029	0.7863±0.0143
	North America	0.3174±0.0048	0.1886±0.0034	0.8995±0.0042
	South America	0.2896±0.0167	0.1694±0.0113	0.8399±0.0023
	Central Asia	0.2501±0.0011	0.1429±0.0007	0.9451±0.0004
	East Asia	0.2535±0.0020	0.1258±0.0012	0.9373±0.0038

Table 6.17: Quantitative results for extreme drought detection on real-world data. (±) denotes the standard deviation for 3 runs.

the driver detection step identifies the related drivers to the events. However, at the cost of a slight decrease in accuracy on extreme events prediction. We found that when we trained on the synthetic

Dataset	ΔF1-score (↓)
Synthetic CERRA Reanalysis	+0.96%
ERA5-Land (EUR-11)	+1.93%

Table 6.18: Ablation study on the driver detection step. The metric is F1-score on extreme drought detection where (ΔF1 $=$ F1$_{\text{(without quantization)}}$ $-$ F1$_{\text{(with quantization)}}$).

and real-world EUR-11 data, the F1-score for detecting extreme events increased only by $\sim 0.96\%$ and $\sim 1.93\%$, respectively, compared to the model with quantization. This verifies that the detected drivers are highly correlated with the extremes.

6.4.5.3 Physical consistency.

In Fig. 6.13, we show the relation between the input reanalysis data, extreme droughts, and identified drivers. For this experiment, we selected pixels with extreme events during summer (weeks 25-38) and visualize the average distribution of drivers with time. The red line at Δt_0 indicates the beginning of the extreme droughts. Z_{score} in the underneath curve represents the deviation from the mean computed from the climatology.

It is expected that evaporation reduces soil moisture, which dries out the soil and vegetation

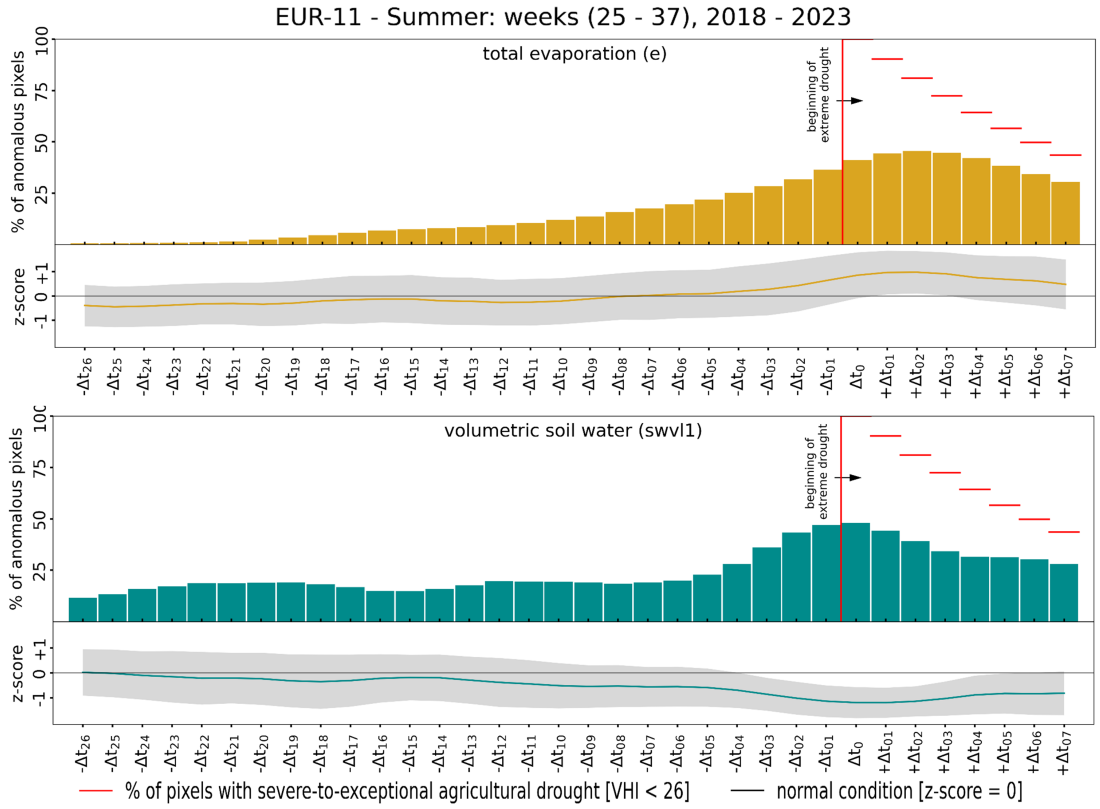


Figure 6.13: Temporal evolution of drivers during the extremes. The results are averaged for the years 2018-2023 to show the overall patterns.

(*Miralles et al.*, 2019). Our model indicates that over Europe, the evaporation and soil moisture are the most informative variables to detect drivers related to extreme droughts. All pixels experienced a pronounced decline in soil moisture and an increase in evaporation as the events evolved. Please see Section 6.4.5.5 for more discussion on the scientific validity.

6.4.5.4 Robustness

To further check the model robustness, we train the model on the same EUR-11 region with 6 different combinations of physical variables. We anticipate that if a variable is relevant to extremes over specific region, e.g., Europe, it should appear in all identified sets of variables, e.g., we expect soil moisture (swvl1) and evaporation (e) to be always presented as explanatory variables for extremes over Europe. Our experimental results in Table 6.19 confirms this assumption.

6.4.5.5 Scientific validity

Surely, soil temperature is a key factor in the drought processing and soil moisture-temperature feedback (*García-García et al.*, 2023). State variables of the land surface such as albedo (fal/al) and soil temperature (stl1/sot) should be very related to reflectance on the ground and consequently to VHI from remote sensing. However, our approach indicates that these variables are not informative enough for the model to identify drivers and predict extremes. To investigate this issue, we conducted

Input variables	Selected variables	F1-score(↑)
{t2m, fal, e, tp, stl1, swvl1}	{e, swvl1}	0.3187 | 0.2152
{d2m, t2m, sp, e, tp, stl1}	{e}	0.3272 | 0.2348
{d2m, fal, sp, e, skt, swvl1}	{e, swvl1}	0.3292 | 0.2241
{t2m, sp, e, skt, stl1, swvl1}	{e, swvl1}	0.3233 | 0.2161
{t2m, fal, sp, skt, stl1, swvl1}	{swvl1}	0.2434 | 0.1605
{t2m, e, tp, swvl1}	{e, swvl1}	0.3038 | 0.2159

Table 6.19: Quantitative results from ERA5-Land EUR-11 data for different combination of physical variables. The metric is F1-score on the extreme drought detection for the validation | test sets.

4 more experiments on both CERRA and ERA5-Land where we trained models that take only one variable al/fal or stl as input and predict the extreme events directly without the driver/anomaly detection step (similar to Section 6.4.5.2). In all of these experiments, the F1-score was very low. In the next experiment, we increased the threshold for VHI and trained new models to predict extremes directly. The results for the validation set are shown below in Table 6.20:

Input variables	Dataset	Domain	VHI < 26	VHI < 40	VHI < 50
{stl1}	ERA5-Land	EUR-11	0.0567	0.3153	0.5836
{t2m, fal, e, tp, stl1, swvl1}	ERA5-Land	EUR-11	0.3380	0.4672	0.6871

Table 6.20: The relation between the definition of extremes from VHI and the model prediction. The metric is F1-score (↑) on the extreme drought detection for the validation sets.

The first potential reason to consider is that some land surface variables might deviate from the reality. ERA5-Land does not use data assimilation directly. The evolution and the simulated land fields are controlled by the ERA5 atmospheric forcing. Another reason might be that when training only on extremes (VHI < 26), there are not enough samples to learn the relations. Note that VHI is a combination of both TCI and VCI. Most extremes (VHI < 26) might result from a deficiency in both stl/t2m and vsw. This might also explain why stl and albedo cannot be that informative to predict very extreme events. Last row in Table 6.19 shows the result when we discard albedo and soil temperature as input variables from ERA5-Land. Moreover, state variables of the hydrological cycle in ERA5-land like volumetric soil water variable (swvl1/vsw) has biases (*Lal et al.*, 2022). One solution to improve the validity of investigation is to use satellite observations for the top layer (*Entekhabi et al.*, 2010). However, the experiments showed that the model relates soil moisture anomalies with the extremes in VHI and provides reasonable predictions.

6.4.5.6 Spatial Resolution

In Table 6.21, we conducted an additional experiment regarding the impact of the spatial resolution. As can be seen the spatial resolution matters as expected.

Dataset	Domain	Spatial resolution	F1-score(↑)
ERA5-Land	EUR-11	0.1°	0.3187
ERA5-Land	EUR-11	0.2°	0.3009

Table 6.21: The impact of spatial resolution on the model prediction. The metric is F1-score on the extreme drought detection for the validation sets.

6.4.5.7 Qualitative results and spatial distribution.

In Fig. 6.14, we show the spatial distribution of the identified drivers at a specific time over EUR-11. Shown are the identified drivers up to 7 weeks (Δt_{-7}) before the extreme agricultural droughts at time Δt_0. We can see that the prediction of drivers and extremes are spatially correlated with the ground truth.

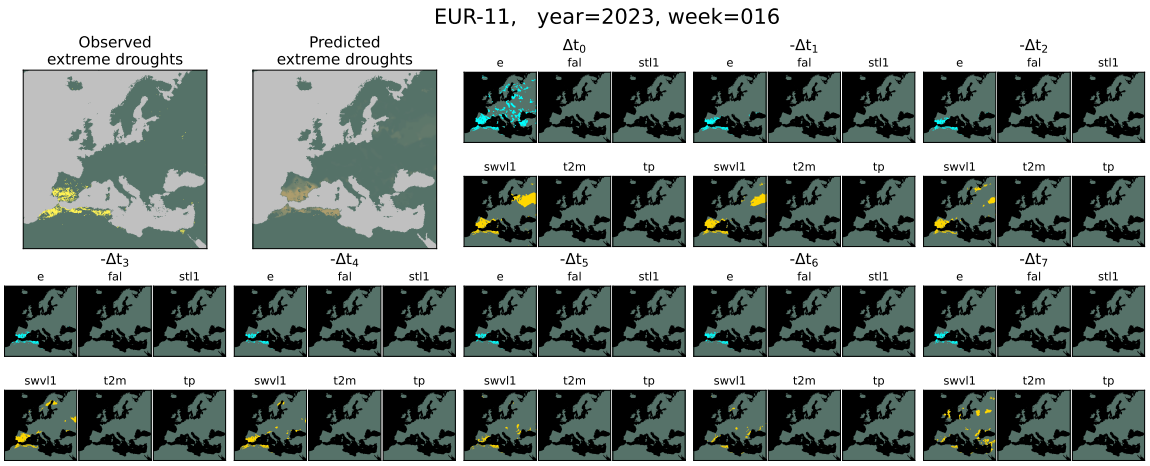


Figure 6.14: Qualitative results on ERA5-Land over the EUR-11 domain. Shown are the identified drivers localized spatio-temporally 7 weeks before the extreme agricultural drought events.

More results are shown in Figs.6.15-6.30. Figs. 6.15 and 6.16 show the averaged spatial distribution of drivers and anomalies related to two regions in Europe. Figs. 6.17 and 6.18 show the qualitative results and temporal evolution of drivers and anomalies related to the extremes on ERA5-Land for Europe (EUR-11). Figs. 6.19 and 6.20 show the qualitative results and temporal evolution of drivers and anomalies related to the extremes on ERA5-Land for Africa (AFR-11). Figs. 6.21 and 6.22 show the qualitative results and temporal evolution of drivers and anomalies related to the extremes on ERA5-Land for South America (SAM-11). Figs. 6.23 and 6.24 show the qualitative results and temporal evolution of drivers and anomalies related to the extremes on ERA5-Land for North America (NAM-11). Figs. 6.25 and 6.26 show the qualitative results and temporal evolution of drivers and anomalies related to the extremes on ERA5-Land for East Asia (EAS-11). Figs. 6.27 and 6.28 show the qualitative results and temporal evolution of drivers and anomalies related to the extremes on ERA5-Land for Central Asia (CAS-11). Figs. 6.29 and 6.30 show the qualitative results and temporal evolution of drivers and anomalies related to the extremes on CERRA reanalysis for Europe. As can be seen, the model can localize extreme droughts for different regions globally.

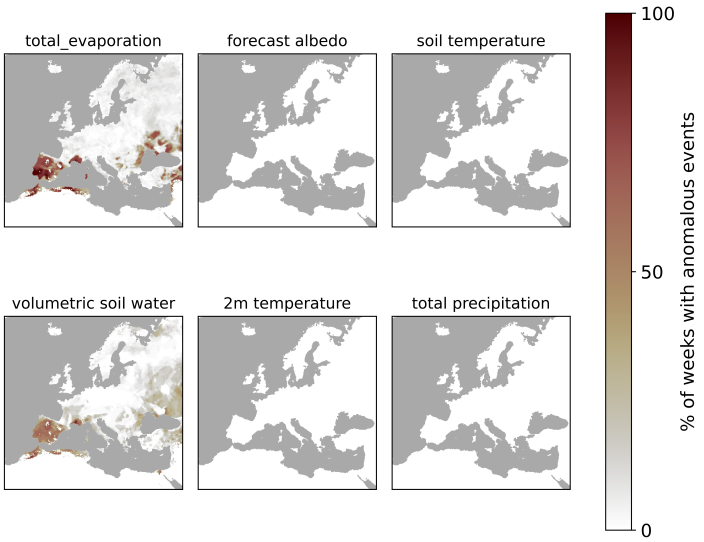


Figure 6.15: The averaged spatial distribution of drivers and anomalies related to Portugal in Europe. For this experiment, we use prediction on EUR-11 from ERA5-Land and select frames (times) within the period 2018-2024 where there were extreme drought of at least 25% of the pixels in the Portugal. Then we normalize the identified drivers and anomalies by the total number of frames to obtain the final map. As can be seen drivers are spatially centered around where extremes were reported.

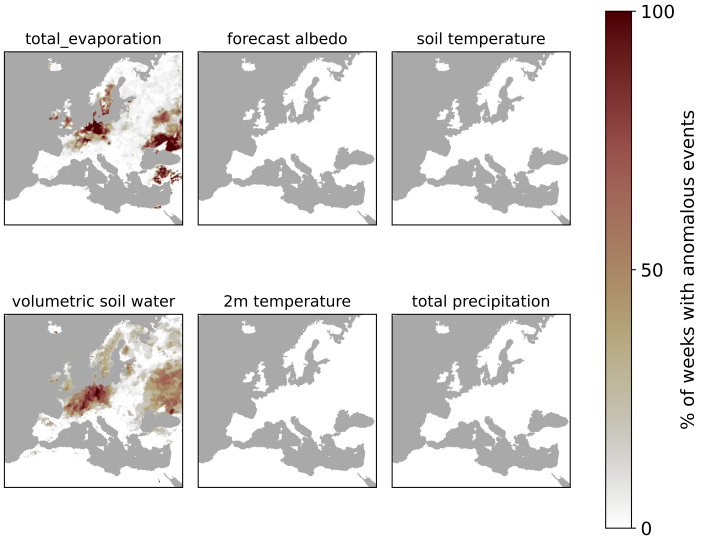


Figure 6.16: The averaged spatial distribution of drivers and anomalies related to a specific place in Europe (North Rhine-Westphalia). For this experiment, we use prediction on EUR-11 from ERA5-Land and select frames (times) within the period 2018-2024 where there were extreme drought of at least 25% of the pixels in the North Rhine-Westphalia. Then we normalize the identified drivers and anomalies by the total number of frames to obtain the final map. As can be seen drivers are spatially centered around where extremes were reported.

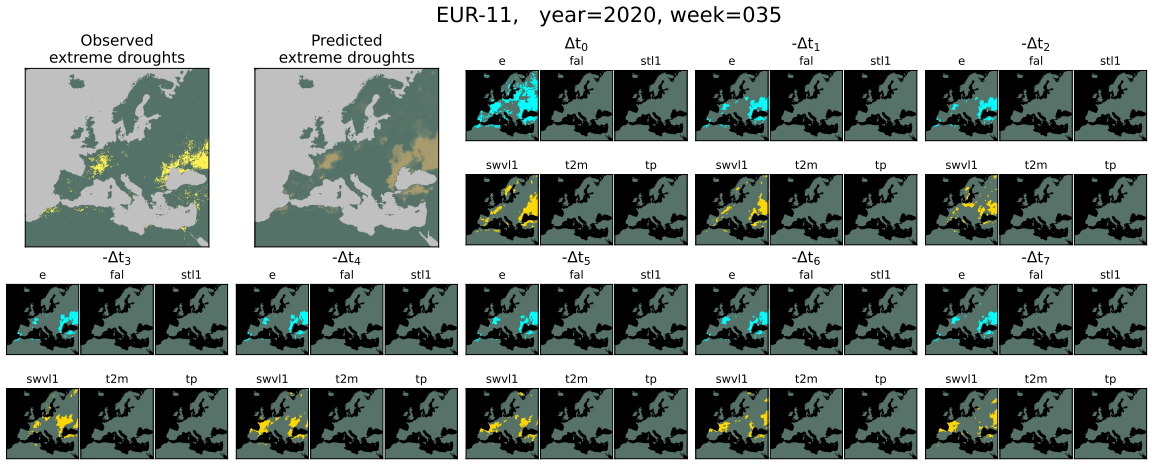


Figure 6.17: Qualitative results on ERA5-Land for Europe (EUR-11). Shown are the identified drivers and anomalies for each variable along with the prediction of extreme agricultural droughts on the top left.

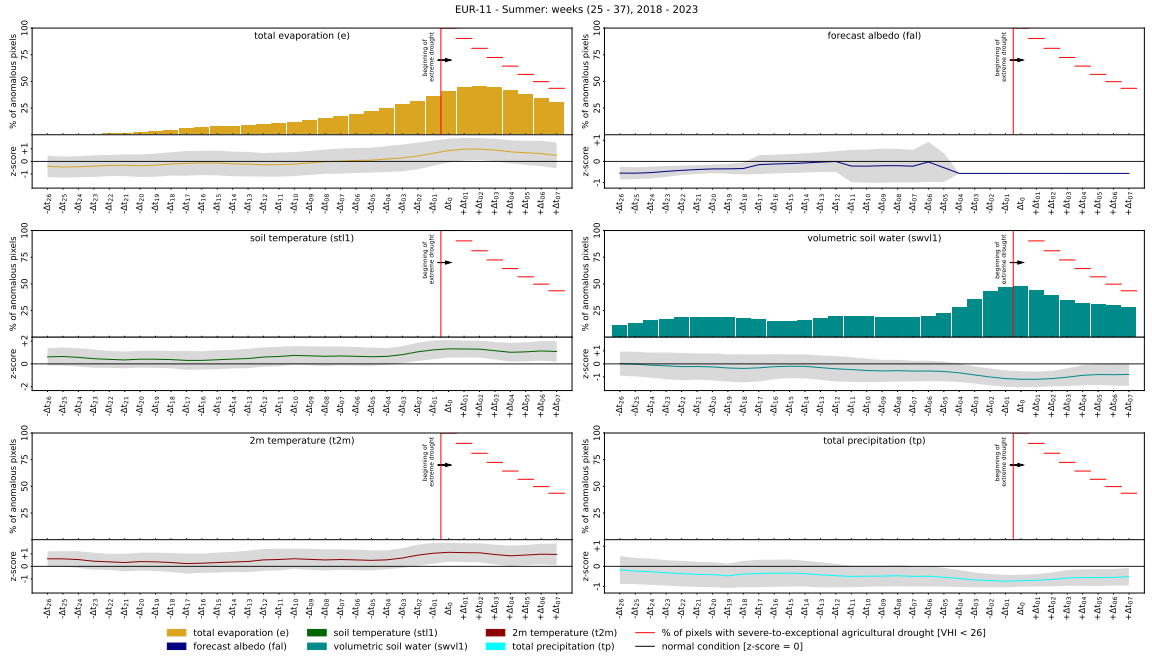


Figure 6.18: Temporal evolution of drivers and anomalies related to the extremes in ERA5-Land for Europe (EUR-11). For this experiments, we select pixels with extreme events during summer (weeks 25-38) for the years 2018-2023 and compute the average distribution of drivers and anomalies with time. The red line at δt_0 indicates the beginning of the extreme droughts. Z_{score} in the underneath curve represents the deviation from the mean computed from the ERA5-Land climatology.

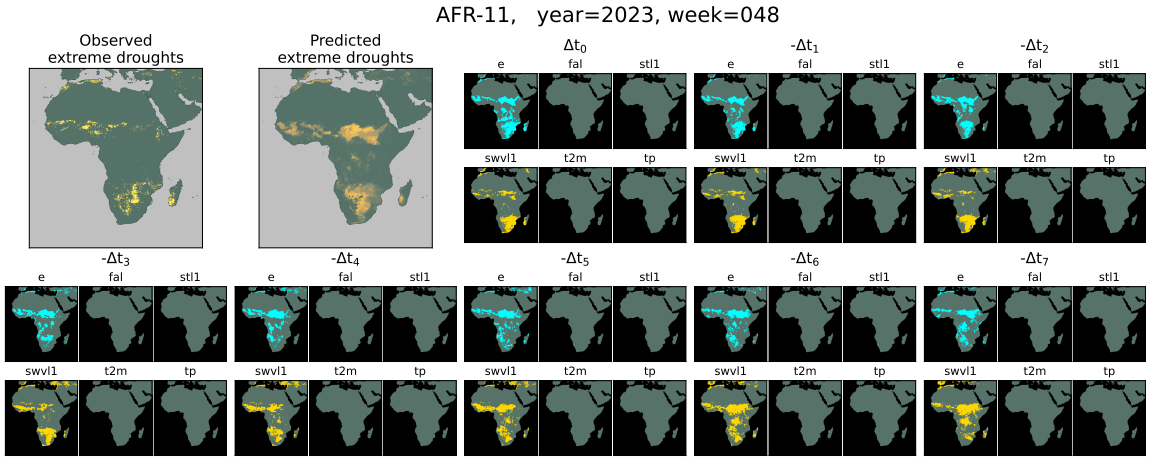


Figure 6.19: Qualitative results on ERA5-Land for Africa (AFR-11). Shown are the identified drivers and anomalies for each variable along with the prediction of extreme agricultural droughts on the top left.

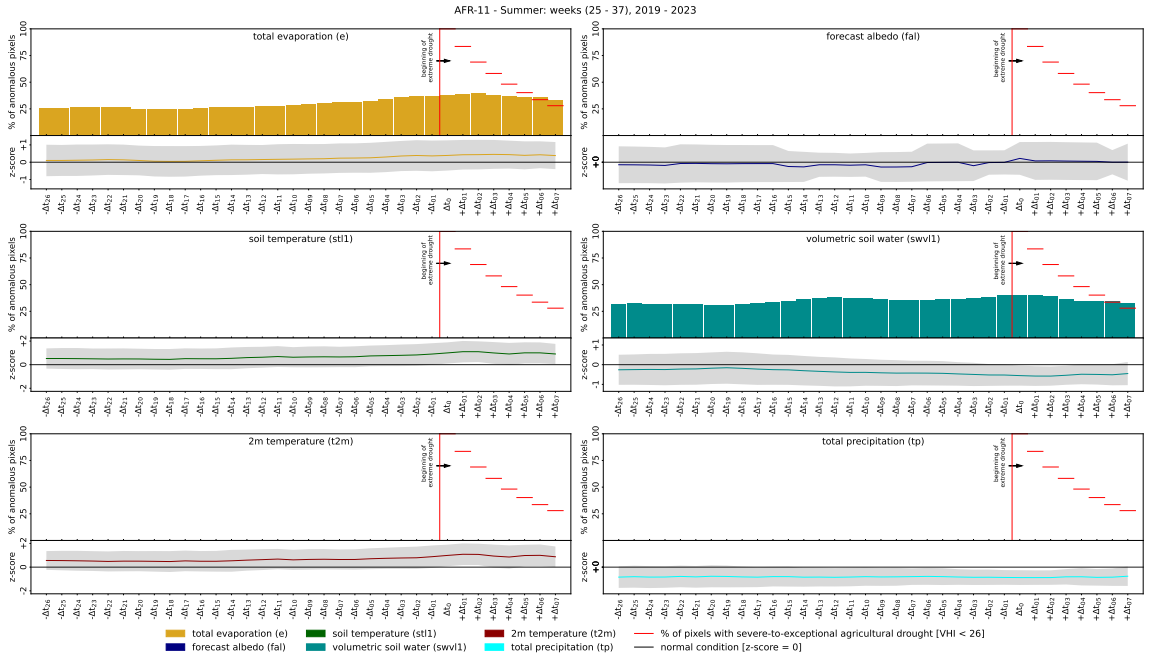


Figure 6.20: Temporal evolution of drivers and anomalies related to the extremes in ERA5-Land for Africa (AFR-11). For this experiments, we select pixels with extreme events during summer (weeks 25-38) for the years 2019-2023 and compute the average distribution of drivers and anomalies with time. The red line at δt_0 indicates the beginning of the extreme droughts. Z_{score} in the underneath curve represents the deviation from the mean computed from the ERA5-Land climatology.

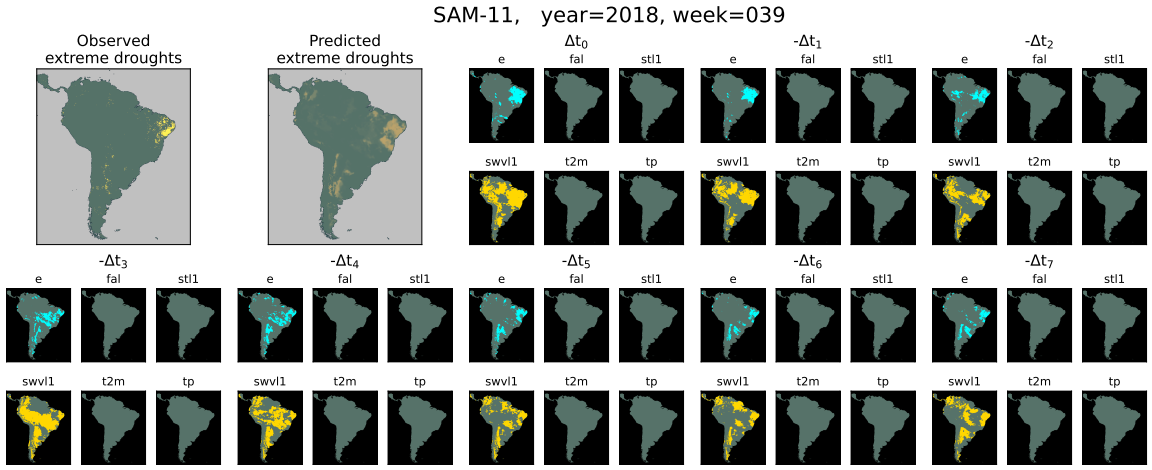


Figure 6.21: Qualitative results on ERA5-Land for South America (SAM-11). Shown are the identified drivers and anomalies for each variable along with the prediction of extreme agricultural droughts on the top left.

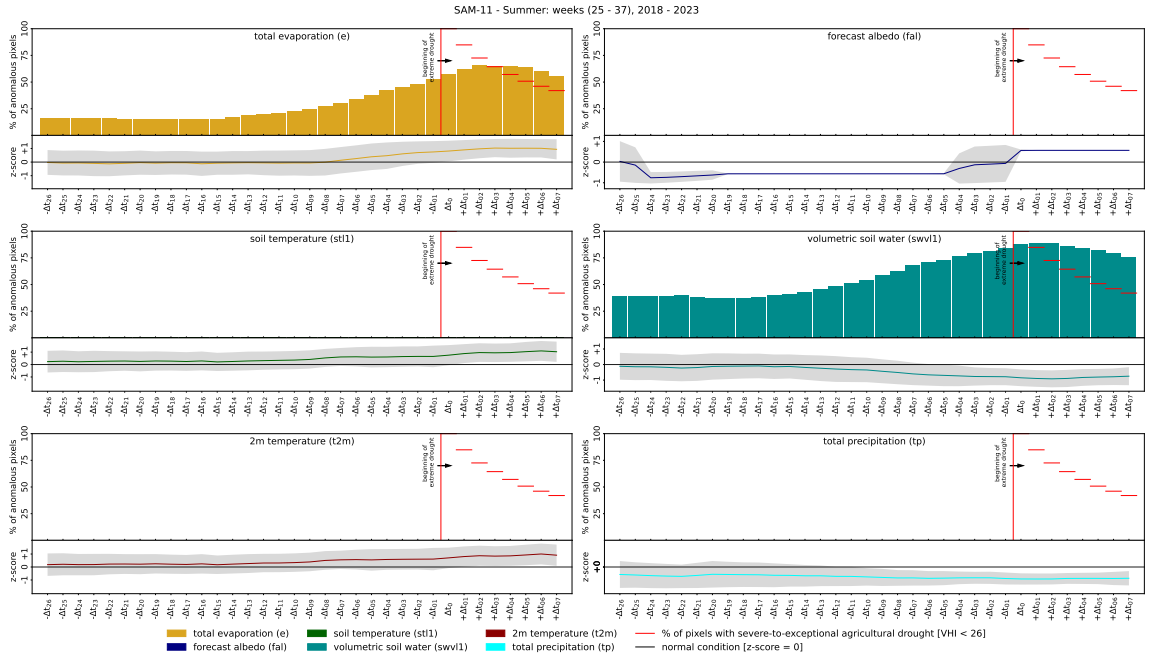


Figure 6.22: Temporal evolution of drivers and anomalies related to the extremes in ERA5-Land for South America (SAM-11). For this experiments, we select pixels with extreme events during summer (weeks 25-38) for the years 2018-2023 and compute the average distribution of drivers and anomalies with time. The red line at δt_0 indicates the beginning of the extreme droughts. Z_{score} in the underneath curve represents the deviation from the mean computed from the ERA5-Land climatology.

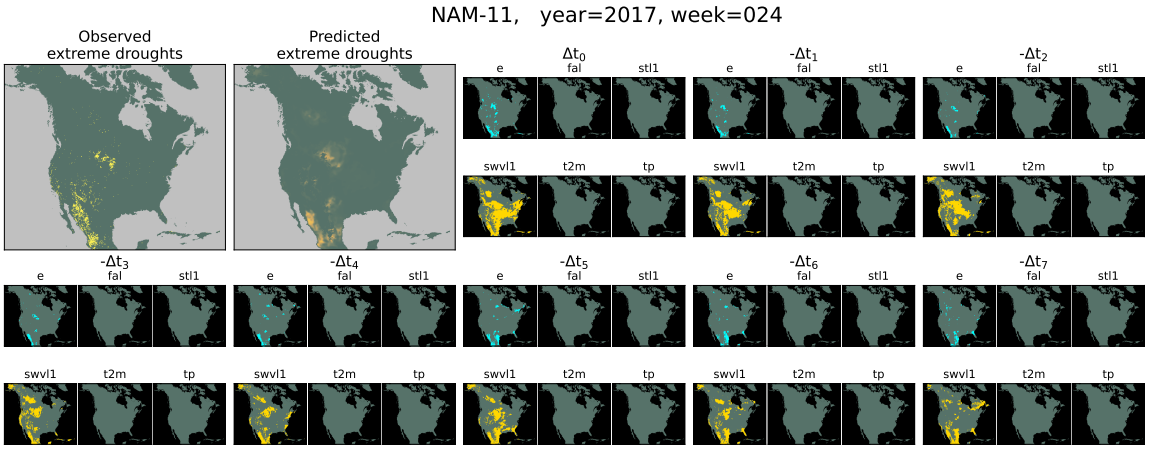


Figure 6.23: Qualitative results on ERA5-Land for North America (NAM-11). Shown are the identified drivers and anomalies for each variable along with the prediction of extreme agricultural droughts on the top left.

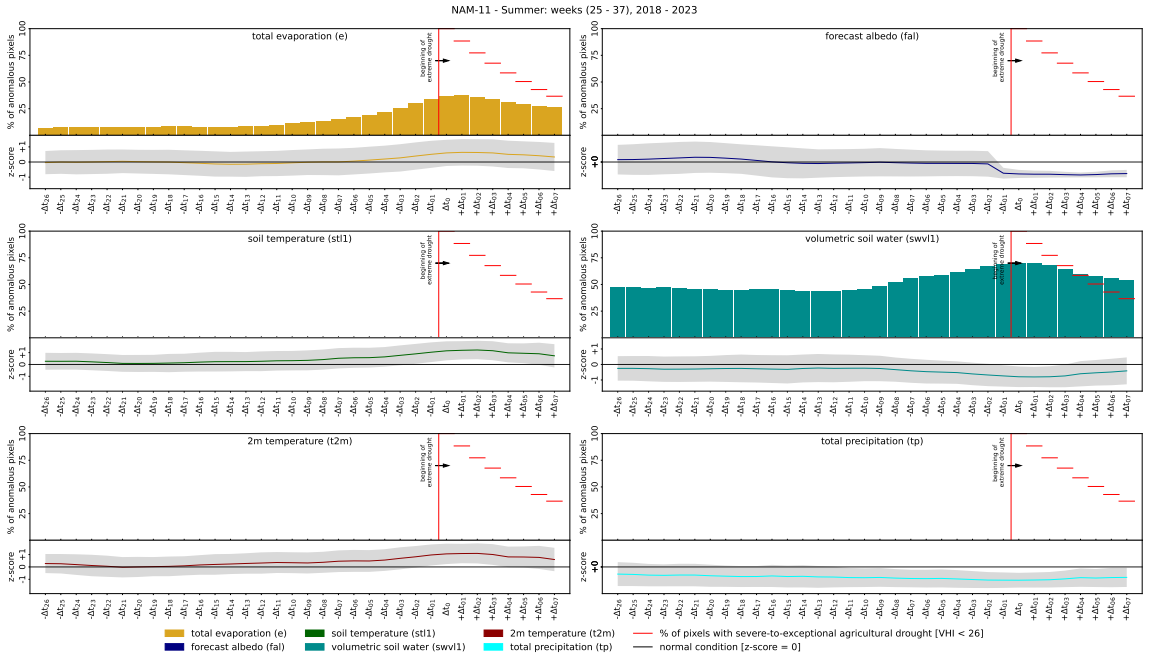


Figure 6.24: Temporal evolution of drivers and anomalies related to the extremes in ERA5-Land for North America (NAM-11). For this experiments, we select pixels with extreme events during summer (weeks 25-38) for the years 2018-2023 and compute the average distribution of drivers and anomalies with time. The red line at δt_0 indicates the beginning of the extreme droughts. Z_{score} in the underneath curve represents the deviation from the mean computed from the ERA5-Land climatology.

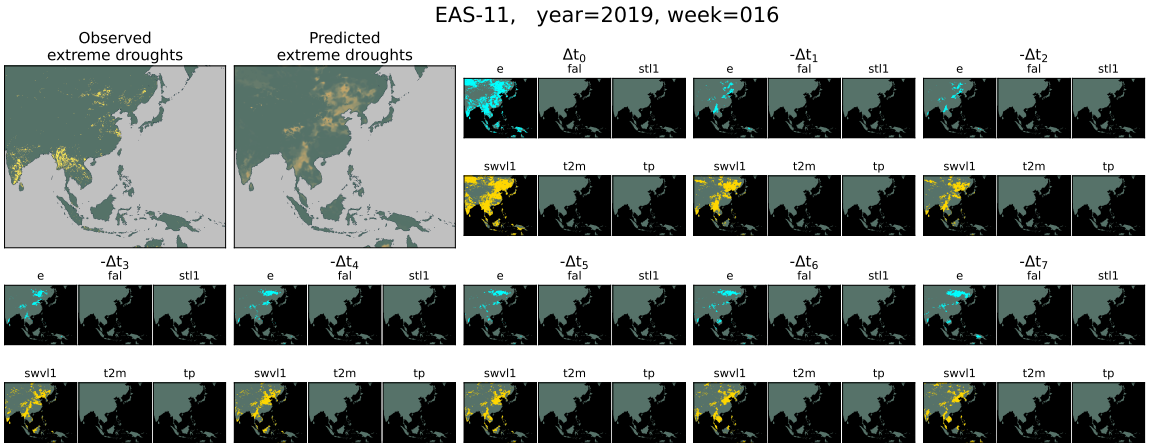


Figure 6.25: Qualitative results on ERA5-Land for East Asia (EAS-11). Shown are the identified drivers and anomalies for each variable along with the prediction of extreme agricultural droughts on the top left.

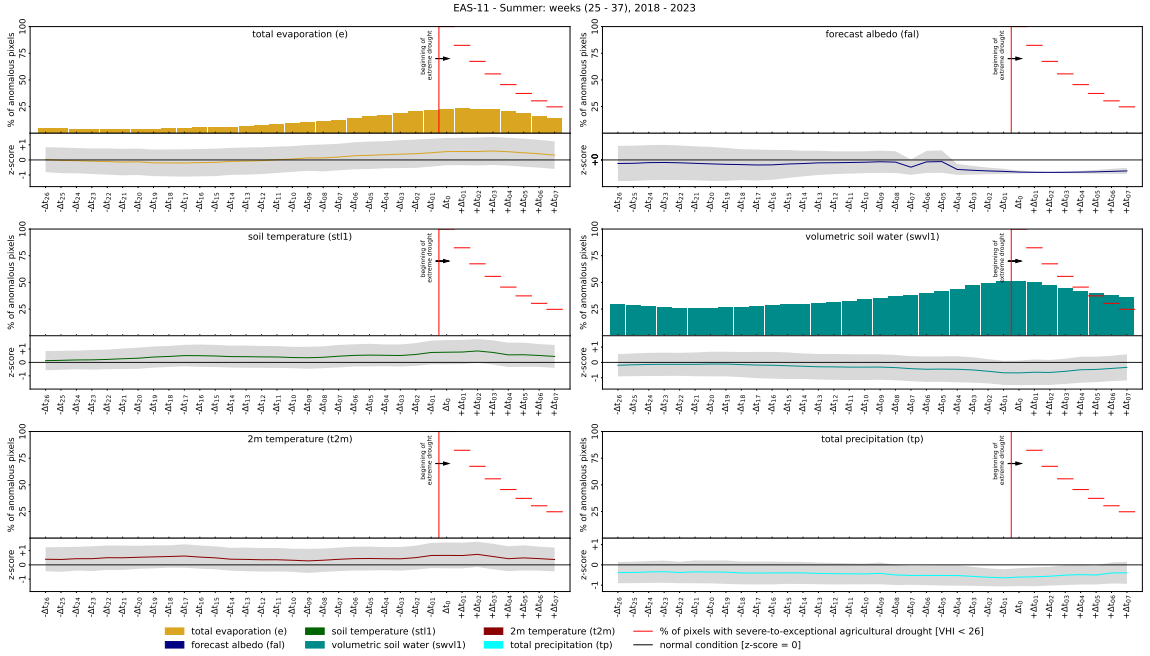


Figure 6.26: Temporal evolution of drivers and anomalies related to the extremes in ERA5-Land for East Asia (EAS-11). For this experiments, we select pixels with extreme events during summer (weeks 25-38) for the years 2018-2023 and compute the average distribution of drivers and anomalies with time. The red line at δt_0 indicates the beginning of the extreme droughts. Z_{score} in the underneath curve represents the deviation from the mean computed from the ERA5-Land climatology.

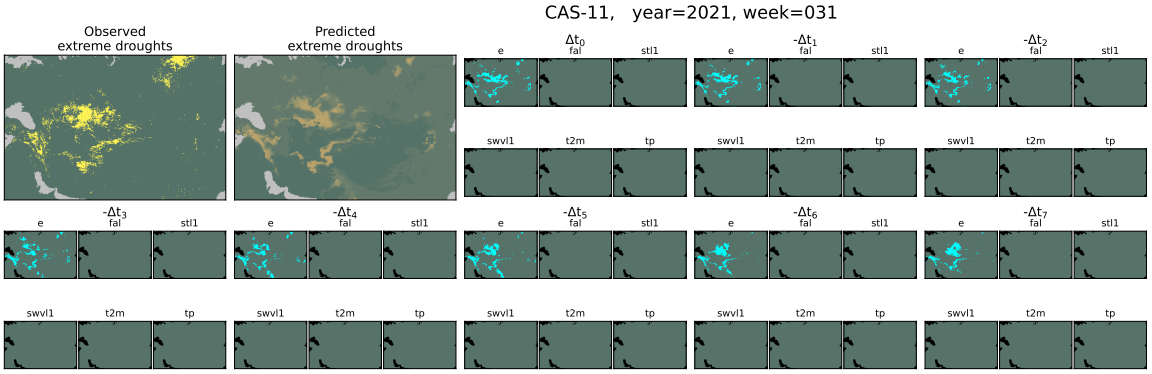


Figure 6.27: Qualitative results on ERA5-Land for Central Asia (CAS-11). Shown are the identified drivers and anomalies for each variable along with the prediction of extreme agricultural droughts on the top left.

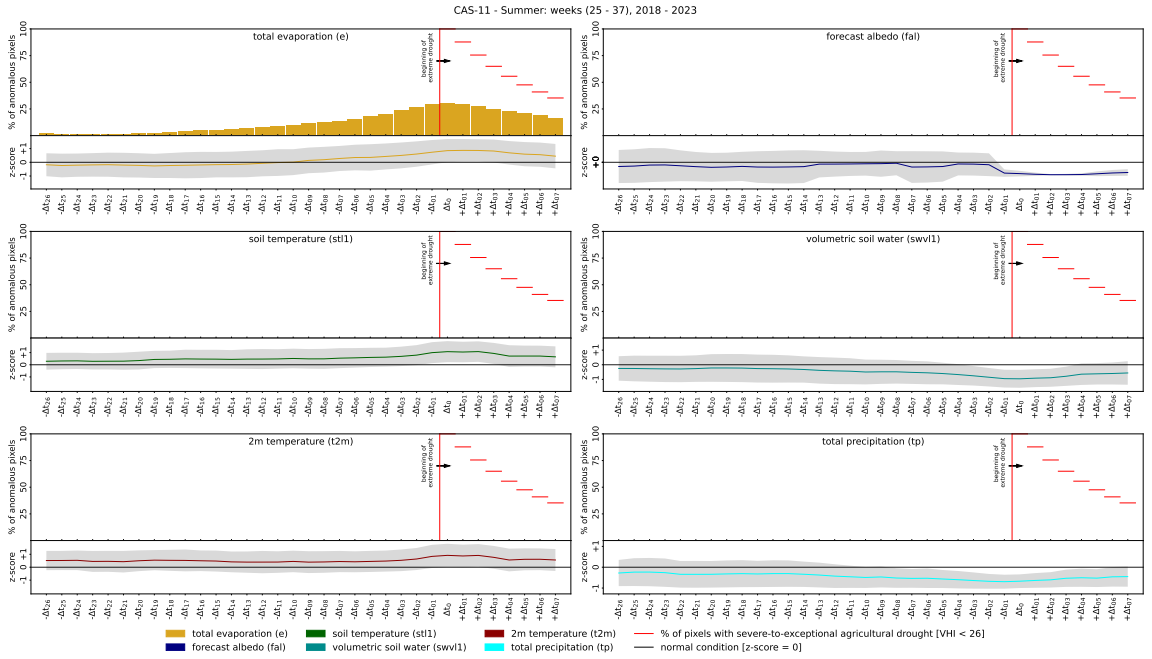


Figure 6.28: Temporal evolution of drivers and anomalies related to the extremes in ERA5-Land for Central Asia (CAS-11). For this experiments, we select pixels with extreme events during summer (weeks 25-38) for the years 2018-2023 nd compute the average distribution of drivers and anomalies with time. The red line at δt_0 indicates the beginning of the extreme droughts. Z_{score} in the underneath curve represents the deviation from the mean computed from the ERA5-Land climatology.

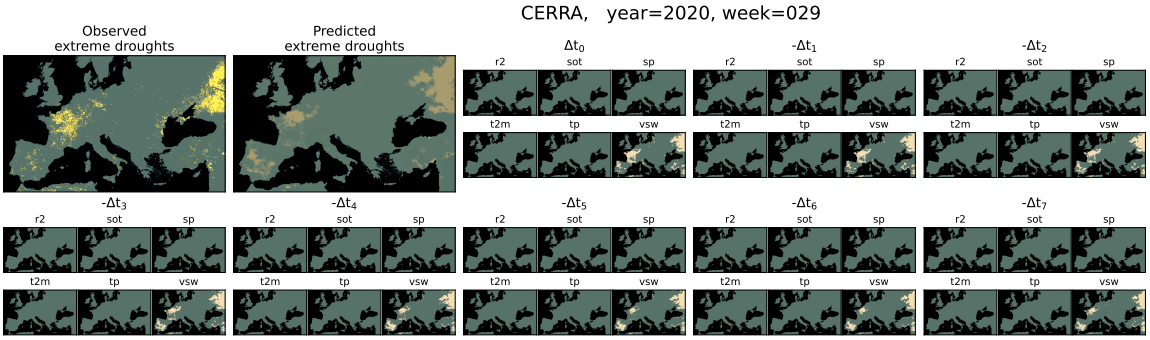


Figure 6.29: Qualitative results on CERRA reanalysis for Europe. Shown are the identified drivers and anomalies for each variable along with the prediction of extreme agricultural droughts on the top left.

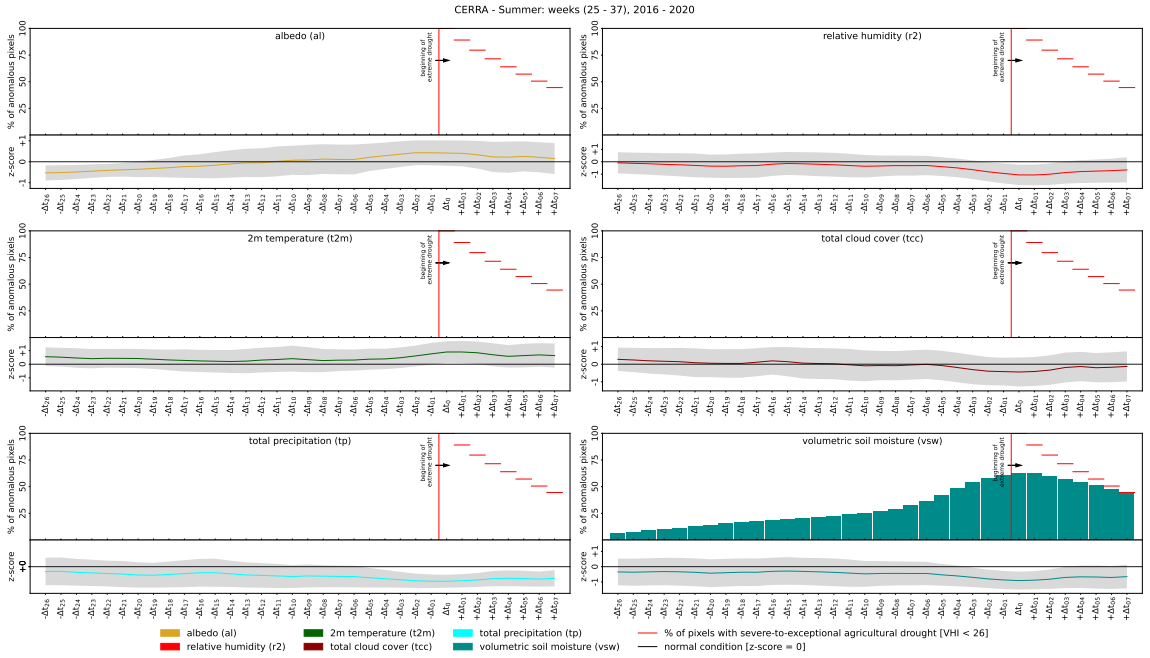


Figure 6.30: Temporal evolution of drivers and anomalies related to the extremes in CERRA reanalysis for Europe. For this experiments, we select pixels with extreme events during summer (weeks 25-38) for the years 2016-2020 and compute the average distribution of drivers and anomalies with time. The red line at δt_0 indicates the beginning of the extreme droughts. Z_{score} in the underneath curve represents the deviation from the mean computed from the CERRA climatology.

6.5 Summary

In this chapter, we introduced a model that can identify the spatio-temporal relations between impacts of extreme events and their drivers. For this, we assumed that there exist precursor drivers, primarily as anomalies in assimilated land surface and atmospheric data, for every observable impact of extremes. We demonstrated the effectiveness of our approach by measuring to which degree the identified drivers can be used to predict extreme agricultural droughts. Apart from experiments on two real-world datasets, we also presented a new framework to generate synthetic datasets that can be used for spatio-temporal anomaly detection and climate research. The results on the synthetic datasets show that the approach is not limited to droughts and can be applied to other extremes. The identified spatio-temporal relations between extreme events and their drivers could support the understanding and forecasting of extremes.

CHAPTER 7

RiverMamba: A State Space Model for Global River Discharge and Flood Forecasting

This chapter is based on the following publication (*Shams Eddin et al.*, 2025a):

RiverMamba: A State Space Model for Global River Discharge and Flood Forecasting
Mohamad Hakam Shams Eddin, Yikui Zhang, Stefan Kollet, and Juergen Gall
Advances in Neural Information Processing Systems 39 (NeurIPS), 2025
Doi: 10.48550/arXiv.2505.22535
Url: https://hakamshams.github.io/RiverMamba

Below we summarize the contributions of each author.

- **Mohamad Hakam Shams Eddin**
  Proposed the initial idea of the paper, which was further refined and extended through joint discussions with Juergen Gall. Hakam designed and developed the methodology of RiverMamba. As the main author, Hakam made contributions to all stages of the project, including designing the experimental setup, conducting the experiments, preprocessing the data, implementation, evaluation, and paper writing.

- **Yikui Zhang**
  Implemented and trained the LSTM model and preprocessed the observational GRDC data. Yikui also contributed significantly in writing and discussions.

- **Stefan Kollet**
  Contributed by supervision, discussions and suggestions.

- **Juergen Gall**
  Juergen proposed the task of flood forecasting and funded the project. He supervised the project development and made substantial contributions through suggestions, discussions and writing.

Contents

7.1 Introduction 154
7.2 Method 156
7.2.1 Hindcast Layer 157

7.2.2 Hindcast Block . . . 158
7.2.3 Serialization . . . 159
7.2.4 Location-aware Adaptive Normalization Layers (LOAN) . . . 160
7.2.5 Mamba Block . . . 160
7.2.6 Forecasting Layer . . . 163
7.2.7 Training . . . 165
7.3 Experimental results . . . 166
7.3.1 Experimental Setup . . . 166
7.3.2 Experiments on Gauged GloFAS River Discharge Reanalysis . . . 171
7.3.3 Experiments on Gauged GRDC Observational River Discharge . . . 181
7.3.4 Comparison with Google Reforecast on Ungauged GRDC . . . 192
7.3.5 Comparison with Google Reforecast on Gauged GRDC . . . 195
7.3.6 Comparison to Operational GloFAS Forecsating on Gauged GRDC . . . 195
7.3.7 Case Studies of Extreme Flood Events . . . 198
7.3.8 Ablation Studies . . . 203
7.4 Summary . . . 209

In Chapters 4 and 5, we focused on forecasting impacts of extreme events and in Chapter 6, we focused on identifying drivers of impacts of extremes. In this chapter, we focus on a specific type of extremes at a global scale, namely, extreme river discharge and floods. For this task, we need a different methodology to handle the sparsity and structure of the river networks. Recent deep learning approaches for river discharge forecasting have improved the accuracy and efficiency in flood forecasting, enabling more reliable early warning systems for risk management. Nevertheless, existing deep learning approaches in hydrology remain largely confined to local-scale applications and do not leverage the inherent spatial connections of bodies of water. Thus, there is a strong need for new deep learning methodologies that are capable of modeling spatio-temporal relations to improve river discharge and flood forecasting for scientific and operational applications. In this chapter, we present RiverMamba, a novel deep learning model that is pretrained with long-term reanalysis data and that can forecast global river discharge and floods on a 0.05° grid up to 7 days lead time, which is of high relevance in early warning. To achieve this, RiverMamba leverages efficient Mamba blocks that enable the model to capture spatio-temporal relations in very large river networks and enhance its forecast capability for longer lead times. The forecast blocks integrate ECMWF HRES meteorological forecasts, while accounting for their inaccuracies through spatio-temporal modeling. We evaluate RiverMamba in this chapter on both long-term real-world reanalysis and observational data. In addition, we demonstrate with analyses that RiverMamba provides reliable predictions of river discharge across various flood return periods, including extreme floods, and lead times, surpassing both AI- and physics-based models.

7.1 Introduction

Riverine floods are one of the most destructive natural disasters, with their risk anticipated to rise in the future as a result of climate change and socioeconomic developments (*Dottori et al.*, 2018; *Merz*

et al., 2021; *Kreibich et al.*, 2022; *Rentschler et al.*, 2022; *Lahsen and Ribot*, 2022). They arise from compound effects, including atmospheric conditions like heavy precipitation caused by circulation patterns and snowmelt succeeding high temperature, all shaped by the specific characteristics of the river drainage area (*Jiang et al.*, 2022a). The interaction of these elements influences flood timing, scale, and severity (*Jiang et al.*, 2022a). This complexity makes future flood risk assessment challenging, as a changing climate may alter these drivers in unpredictable ways (*Jiang et al.*, 2024). Therefore, early prediction of flood risk, especially for extreme floods, is a key measure for effective flood risk mitigation (*Pappenberger et al.*, 2015; *Camps-Valls et al.*, 2025).

To support national forecasting initiatives, current operational flood early warning systems can forecast river discharge in real-time and provide flood forecasts at different scales (*Emerton et al.*, 2016; *Hapuarachchi et al.*, 2022; *Dottori et al.*, 2017). The discharge forecasts derived from these systems can be further processed using inundation models to create anticipated flooded areas (*Nevo et al.*, 2022; *Najafi et al.*, 2024). The Global Flood Awareness System (GloFAS) (*Alfieri et al.*, 2013; *Harrigan et al.*, 2023), developed under the Copernicus Emergency Management Service (CEMS) and operated by the European Centre for Medium-Range Weather Forecasts (ECMWF), represents the cutting-edge physics-based model for real-time and worldwide hydrological forecasting. However, physics-based hydrological models are expensive to run and require extensive calibration to handle complex catchment characteristics.

AI-based early warning systems are thus considered as vital tools to enhance climate risk resilience (*Jones et al.*, 2023; *Reichstein et al.*, 2025) and to enable flood forecasting without requiring full physical process understanding (*Nearing et al.*, 2021; *Frame et al.*, 2022). While deep-learning approaches for weather forecasting (*Pathak et al.*, 2022; *Price et al.*, 2025; *Bi et al.*, 2023) have been investigated in recent years, very little work has been done for forecasting river discharge at large spatial regions since it is very challenging. It requires the combination of sparse gauged river observations with high-resolution land surface, re-analysis, and weather forecast data. Furthermore, floods occur rarely and the goal is to forecast floods of different severity as shown in Fig. 7.1. Recently, an LSTM-based model has been proposed *Nearing et al.* (2024). While it achieves promising results, it forecasts floods only locally at sparse river basins and does not consider routing. Modeling spatio-temporal relations, however, is very important and required to generate consistent dense maps as in Fig. 7.1, since river discharge at points near connected bodies of water is highly correlated.

In this chapter, we propose the first deep learning approach for global river discharge and flood forecasting that is not only capable of forecasting at sparse gauged observation points, but also of forecasting accurate, high-resolution (0.05°) global river discharge maps. In order to deal with the sparseness of gauged river points and the computational complexity of modeling spatio-temporal relations at the global scale, our proposed RiverMamba leverages Mamba blocks, which are bidirectional state space models (*Gu et al.*, 2021, 2022; *Smith et al.*, 2023; *Gu and Dao*, 2024), and spatio-temporal forecast blocks. Using a specialized procedure to convert sampled points into 1D sequences, RiverMamba maintains a very large spatio-temporal receptive field, connecting the routing of the river channel networks and the teleconnection of meteorological data across space and time. RiverMamba has thus the possibility to consider a spatio-temporal context that covers very large river networks like the Amazon River. The forecast layers are further forced by high-resolution meteorological data (HRES) to generate medium-range river discharge forecasts up to 7 days lead time. To address uncertainty in the meteorological forcing, we built the forecast layers so that they can, for each catchment point, incorporate information about meteorological forcing from the neighbor-

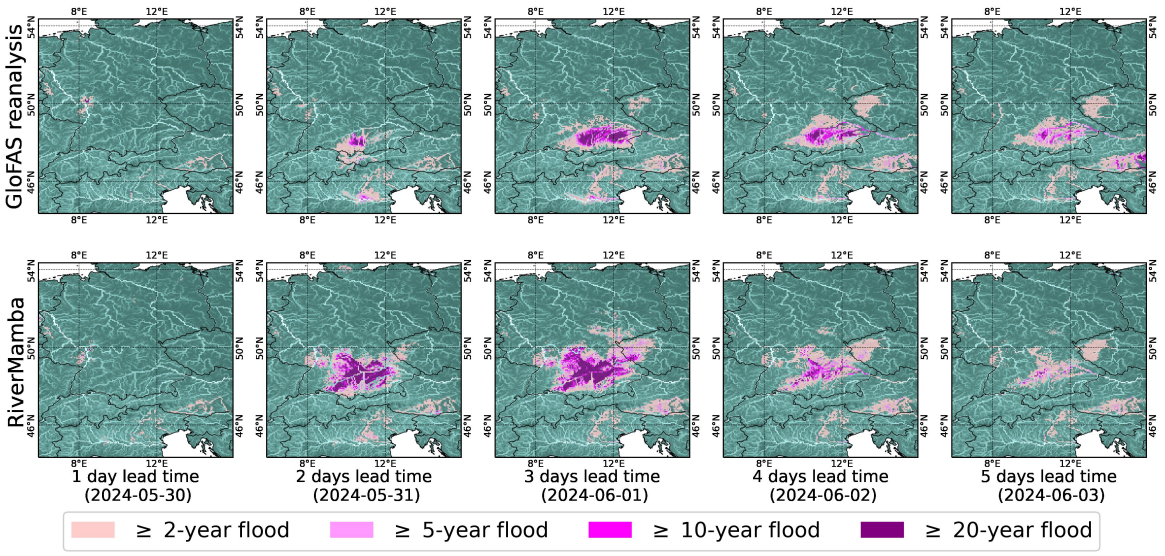


Figure 7.1: Example of a 5-day forecast of river discharge and flood events. In early June 2024, a significant flood affected Southern Germany. While the top row shows the floods obtained from the GloFAS reanalysis, the bottom row shows the river discharge forecast by our approach. The severity of floods is categorized by the statistical flood return period, e.g., occurring every 10 years.

ing points and throughout the temporal dimension. Thus, RiverMamba ensures a consistent forecast through space and time.

7.2 Method

In this chapter, we present the first deep learning approach that not only forecasts flood events at sparse gauged river observations, but that is capable of forecasting accurate, high-resolution (e.g., at 0.05°) maps of river discharge up to few days at global scale, as shown in Figs. 7.1 and 7.2. These maps are essential to forecast flood events of various severity like a flood that re-occurs statistically within a 1.5-year return period or a 'flood of the century'. This is very challenging since it requires a model that models spatial-temporal relations in an efficient way and integrates different sources of data (Fig. 7.2).

As input, we use the initial condition of the forecasts from ERA5-Land reanalysis (Section 2.2.4), denoted by $\mathbf{X}_{ERA5}^{t-T:t-1} = \{\mathbf{X}_{ERA5}^{t-T}, \ldots, \mathbf{X}_{ERA5}^{t-2}, \mathbf{X}_{ERA5}^{t-1}\}$, the initial condition $\mathbf{X}_{GloFAS}^{t-T:t-1}$ from the GloFAS reanaylsis data (Section 2.2.6), and the initial condition $\mathbf{X}_{CPC}^{t-T-1:t-2}$ from the operational global unified gauge-based analysis of daily precipitation (Section 2.2.9). We also include data from weather forecasts, where we use the high-resolution meteorological forcing forecasts $\mathbf{X}_{HRES}^{t+1:t+L}$ from the ECMWF Integrated Forecast System (IFS) (Section 2.2.8), where L is the lead time for the forecast. We generate the river discharge forecast at t, using 00:00 UTC as reference time, for $t+1$ until $t+L$. This means that we do not address nowcasting but only forecasting as it is more relevant. We also do not include any nowcasts ($\mathbf{X}_{HRES}^{t}$) as input. The rationale behind this is to ensure broader applicability, since many weather forecast systems especially ML models provide forecasts at $t > 0$. However, adding nowcasts to the model is straightforward if they are available. To make the setup

as realistic as possible, we do not include any data after 00:00 UTC and we consider $\mathbf{X}_{GloFAS}$ and $\mathbf{X}_{ERA5}$ at day $t-1$ and $\mathbf{X}_{CPC}$ at day $t-2$. Additionally, we include river attributes $\mathbf{X}_{static}$ like catchment morphology from LISFLOOD (Section 2.2.10). The input variables are described in details in Tables 2.8, 2.12 and 2.13.

In addition, we add the Cartesian coordinates for the points on the WGS-84 ellipsoid to enhance the positional encoding:

$$x = (N + h)\cos\phi\cos\lambda, \quad y = (N + h)\cos\phi\sin\lambda, \quad z = ((1 - e^2) + h)\sin\phi\,, \tag{7.1}$$

$$N = \frac{a}{\sqrt{(1 - e^2\sin^2(\phi))}}, \quad e^2 = \frac{a^2 - b^2}{a^2}\,, \tag{7.2}$$

where N is the radius of curvature in the prime vertical, h is the height from the elevation model, ϕ and λ are the geographic latitude and longitude, respectively, a and b are the semi-major and semi-minor axes of the ellipse, and e is the eccentricity. We set $a = 6,378.137$ km and $b = 6,356.752$ km. The 3D coordinates are added to the static features $\mathbf{X}_{static}$.

Given these inputs, RiverMamba forecasts changes of the daily mean river discharge $\Delta\mathbf{X}_{dis24}^{t+1:t+L}$ relative to the daily mean river discharge at $t-1$, e.g., $\mathbf{X}_{dis24}^{t-1}$. The forecast daily mean river discharge is thus given by $\mathbf{X}_{dis24}^{t+l} = \mathbf{X}_{dis24}^{t-1} + \Delta\mathbf{X}_{dis24}^{t+l}$.

In this work, we formulate this as follows:

$$\begin{aligned} f_\theta : (\{\{\mathbf{X}_{ERA5}^{t-T}, \ldots, \mathbf{X}_{ERA5}^{t-2}, \mathbf{X}_{ERA5}^{t-1}\}; \{\mathbf{X}_{CPC}^{t-T-1}, \ldots, \mathbf{X}_{CPC}^{t-3}, \mathbf{X}_{CPC}^{t-2}\}; \\ \{\mathbf{X}_{GloFAS}^{t-T}, \ldots, \mathbf{X}_{GloFAS}^{t-2}, \mathbf{X}_{GloFAS}^{t-1}\}\}; \{\mathbf{X}_{static}\}; \\ \{\mathbf{X}_{HRES}^{t+1}, \mathbf{X}_{HRES}^{t+2}, \ldots, \mathbf{X}_{HRES}^{t+L}\}) \rightarrow (\{\Delta\mathbf{X}_{dis24}^{t+1}, \Delta\mathbf{X}_{dis24}^{t+2}, \ldots, \Delta\mathbf{X}_{dis24}^{t+L}\})\,, \end{aligned} \tag{7.3}$$

where f is the mapping function parameterized as a neural network.

An overview of RiverMamba is shown in Fig. 7.2. For training, we sample P points that are on the land surface and near water bodies. The details are described in Section 2.2.12. For each point p, we obtain a temporal sequence of embedding vectors $\mathbf{X}_{embed}^{t-T:t-1}(p)$:

$$\mathbf{X}_{embed}^{t}(p) = \mathrm{LN}\Big(\mathrm{Tanh}\Big(\mathrm{Concat}\big(\mathrm{Linear}(\mathbf{X}_{ERA5}^{t}(p)), \mathrm{Linear}(\mathbf{X}_{GloFAS}^{t}(p)), \mathrm{Linear}(\mathbf{X}_{CPC}^{t-1}(p))\big)\Big)\Big)\,, \tag{7.4}$$

where LN is the layer norm and Linear is the projection layer. The dimensions of the input are $\mathbf{X}_{ERA5} \in \mathbb{R}^{N\times T\times P\times V_e}$, $\mathbf{X}_{GloFAS} \in \mathbb{R}^{N\times T\times P\times V_g}$, and $\mathbf{X}_{CPC} \in \mathbb{R}^{N\times T\times P\times 1}$, where N is the batch size, V_e is the number of variables from ERA5, and V_g is the number of variables from GloFAS. The embedding $\mathbf{X}_{embed} \in \mathbb{R}^{N\times T\times P\times K}$, where $K = 192$ is the dimensionality of the embedding, is then the input to the encoder defined by the hindcast layers.

7.2.1 Hindcast Layer

The hindcast layers model spatio-temporal relations and aggregate the observations over time. Except for the 1^{st} layer which processes the full temporal resolution, the temporal resolution is downsampled by a factor of 2 with a linear layer at the beginning of each hindcast layer, such that the output of the last hindcast layer $\mathbf{X}_{hindcast} \in \mathbb{R}^{B\times 1\times P\times K}$ has a temporal resolution of $T = 1$. In

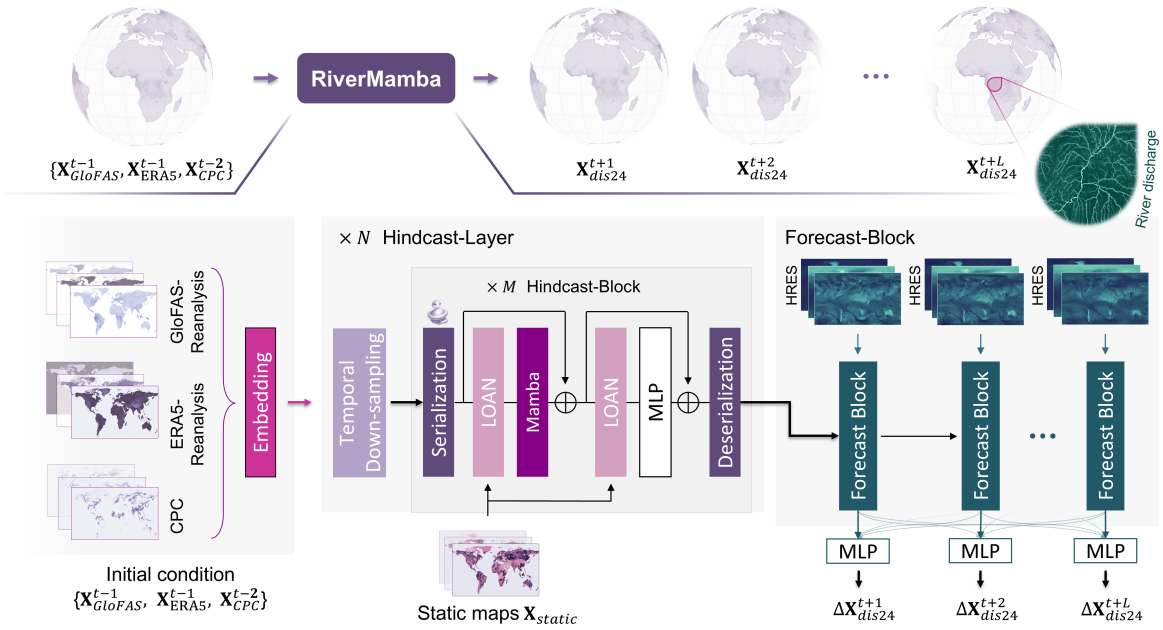


Figure 7.2: An overview of the proposed RiverMamba model for river discharge forecasting. The model forecasts at time t, high-resolution river discharge maps $\mathbf{X}_{dis24}^{t+1:t+L}$ from initial conditions ($\mathbf{X}_{ERA5}^{t-T:t-1}$, $\mathbf{X}_{GloFAS}^{t-T:t-1}$, $\mathbf{X}_{CPC}^{t-T-1:t-2}$), static river attributes ($\mathbf{X}_{static}$), and meteorological forecasts ($\mathbf{X}_{HRES}^{t+1:t+L}$).

our implementation, we chose $T = 4$ as for the GloFAS operational system and a temporal down-sampling of 2. Consequently, we defined 3 layers to encode the input.

The hindcast layers further integrate the static river attributes $\mathbf{X}_{static}$ that contain additional information like catchment morphology, which is relevant for flood forecasting. While we analyze the impact of the different inputs, in particular the river attributes, in Section 7.3.8.5, another key aspect of the hindcast blocks is the specialized serialization of the spatio-temporal points and the Mamba blocks (*Gu and Dao*, 2024; *Liu et al.*, 2024; *Zhu et al.*, 2024). The serialization defines the way the sampled points are connected, and the Mamba block efficiently updates the features of each point based on the spatio-temporal structure. This is a very important design choice since transformer blocks are computationally infeasible for global flood forecasting, whereas *Nearing et al.* (2024) does not consider spatial relations at all.

The output of the last hindcast layer is then processed along with the HRES meteorological forcing by forecast blocks, and MLP-based regression heads predict for each lead time l the difference of daily mean river discharge $\Delta\mathbf{X}_{dis24}^{t+l}$ with respect to the daily mean river discharge at $t-1$. In the following, we describe the components of RiverMamba in details.

7.2.2 Hindcast Block

As shown in Fig. 7.2, the hindcast block has three main components: serialization and deserialization, location-aware adaptive normalization layers (LOAN) to integrate static river attributes, and the Mamba block.

7.2.3 Serialization

The serialization defines the spatio-temporal scanning path over all sampled points for the following Mamba block. For this, we propose space-filling curves that sequentially traverse through all points. The concept was introduced in *Peano and Peano* (1990) and the space-filling can be defined as a bijective function $\Omega : \mathbb{Z}^3 \rightarrow \mathbb{N}$, where every point in the discrete space corresponds to a unique index within the sequence. We call this mapping the serialized encoding. Each point is stored as a 64-bit integer. Due to the nature of the bijective transformation, there is an inverse mapping $\Omega^{-1} : \mathbb{N} \rightarrow \mathbb{Z}^3$ which allows for the mapping of the encoded index back into the point's position $p_i \in \mathbb{Z}^3$ (or $p_i \in \mathbb{R}^3$ in case of a continuous space). This inverse mapping is called the serialized decoding or the deserialization. In this chapter, we investigate three curves; the Generalized Hilbert (Gilbert) curve, which is a generalized version of the Hilbert curve (*Hilbert and Hilbert*, 1935), as well as the Sweep and Zigzag curves in vertical and horizontal directions In the following, we describe the mapping for each curve:

Sweep. This curve fills in the space like a spherical helix or a Luxodrome around the sphere.

Zigzag. This curve is similar to the Sweep curve. The main difference is that the transformation ensures that every neighboring points on the curve are also neighboring in the physical space.

Generalized Hilbert. Generalized Hilbert (Gilbert) is a Hilbert space-filling curve (*Hilbert and Hilbert*, 1935) for rectangular domains of arbitrary non-power of two sizes (*Zhang et al.*, 2006). We used the numpy implementation of (`https://github.com/jakubcerveny/gilbert`) to generate the curves. Transposed Gilbert is generated as $y_{(transpose)} = H - y$, where $y \in [1, H]$.

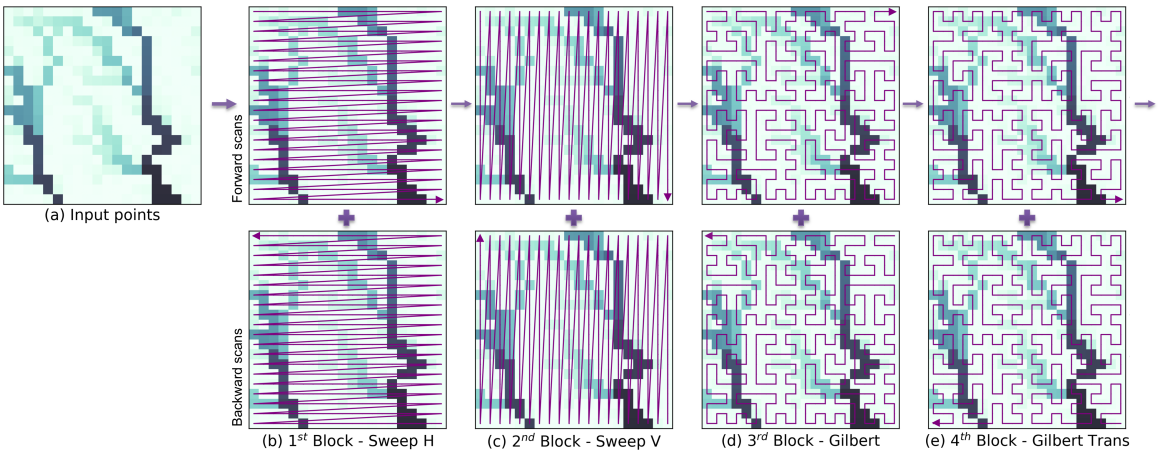


Figure 7.3: Illustration of the spatial scans in RiverMamba.

Examples of space-filling curves in 2D are illustrated in Fig. 7.3. For simplicity, we define the curves on the 2D PlateCarree projection of the Earth. As illustrated in Figs. 7.5 and 7.6, the serialization is done according to the sorted serialized encoding Ω of all points.

As shown in the experimental results (Table 7.16), a combination of Sweep and Gilbert curves performs best. To this end, each hindcast block has its own curve. As shown in Fig. 7.3, we sweep

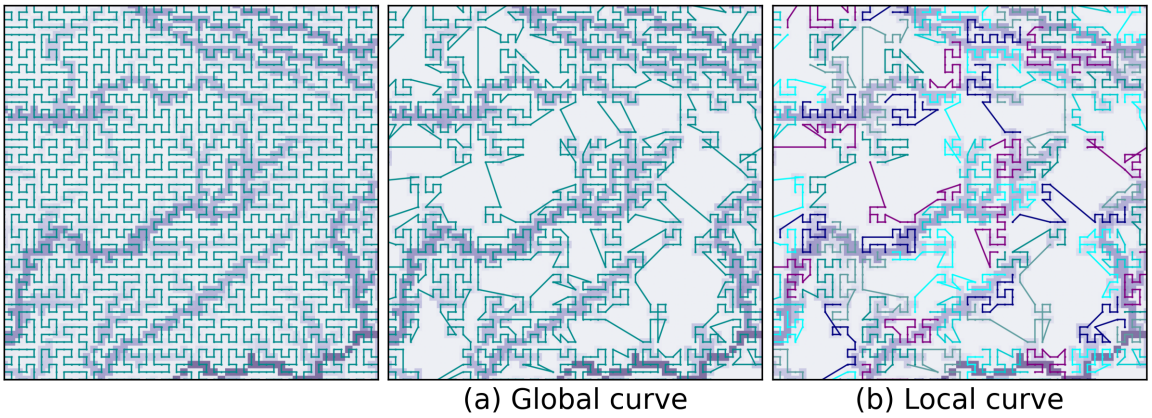


Figure 7.4: Illustration of the difference between global curve (middle) and grouped local curves (right). The left image shows a Gilbert space-filing curve for all points. In our experiments, the global curve is used always (middle). For the experiments with Flash-Attention, we use the local curves (right).

in the first block over the horizontal direction. The spatial curves are connected over time by continuing the last point of the curve at t with the first point of the curve at $t+1$. The second block then sweeps over the vertical direction and we continue with the Gilbert curve and its transposed. These four space-filling curves are iterated. By altering the curves sequentially through the hindcast blocks, the sampled points will be connected and scanned from diverse spatial perspectives, enabling RiverMamba to capture different contextual features.

For our experiments with Flash-Attention, we split the space-filling curve and do a local modeling (see Fig.7.4 (b)), since Flash-Attention does not scale with global data.

7.2.4 Location-aware Adaptive Normalization Layers (LOAN)

In order to condition the model on static river attributes $\mathbf{X}_{static}$, the location-aware adaptive normalization layer (LOAN) (Chapter 4) modulates the features $\mathbf{X}$ within the hindcast block:

$$\text{LOAN}(\mathbf{X}) = \left(\frac{\mathbf{X} - \mu}{\sigma}\right) + \text{GELU}(\text{Linear}(\mathbf{X}_{static}))\,, \tag{7.5}$$

where a linear layer projects $\mathbf{X}_{static} \in \mathbb{R}^{B,1,P,V_s}$, with V_s being the number of static variables, to $\mathbb{R}^{B,1,P,K}$. The output is then duplicated along the temporal dimension so that the output has the dimension $\mathbb{R}^{B,T,P,K}$. $\mu \in \mathbb{R}^{B,T,P,1}$ and $\sigma \in \mathbb{R}^{B,T,P,1}$ are the mean and standard deviation of $\mathbf{X}$ along the channel dimension, respectively, and $\mathbf{X} \in \mathbb{R}^{B,T,P,K}$ is the input to the LOAN layer. Both μ and σ are duplicated K times along the last dimension to match $\mathbf{X}$. The layer normalizes the features and adds a systematic bias based on the attributes. For instance, the features are normalized and biased based on location attributes that have an impact on drainage and floods.

7.2.5 Mamba Block

Fig. 7.7 shows a more detailed structure of the hindcast block with the elements of the Mamba block. After the input is serialized into a 1D sequence based on the block-specific space-filling curve and the

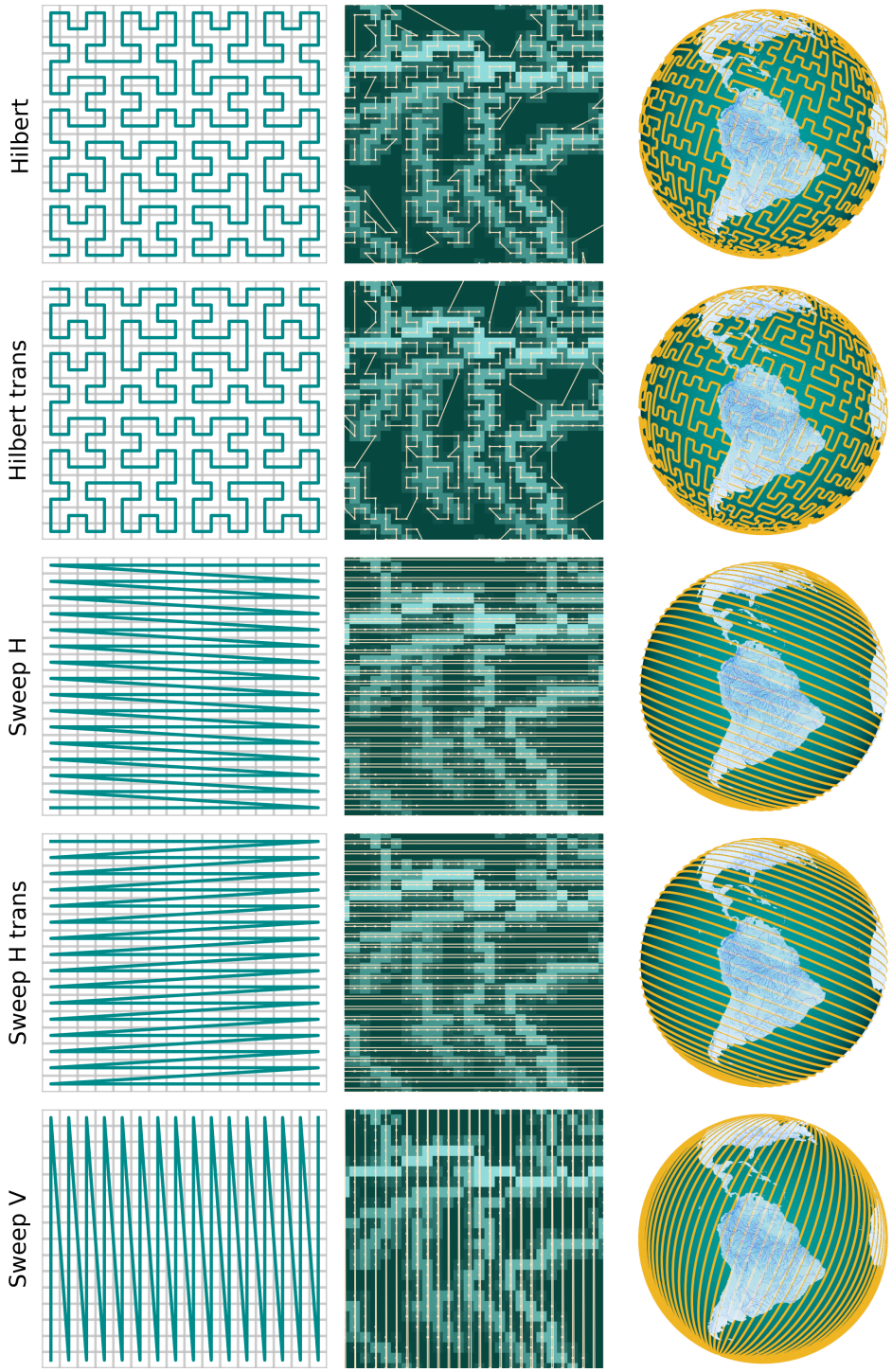


Figure 7.5: Visualization of different types of space-filling curves. For each type, we show the space-filling curve over a 2D discrete space (left), zoomed in version over the Earth where the points are sorted via a specific serialization order within the space-filling curve (middle), and simplified 3D visualization of the curve over the Earth (right).

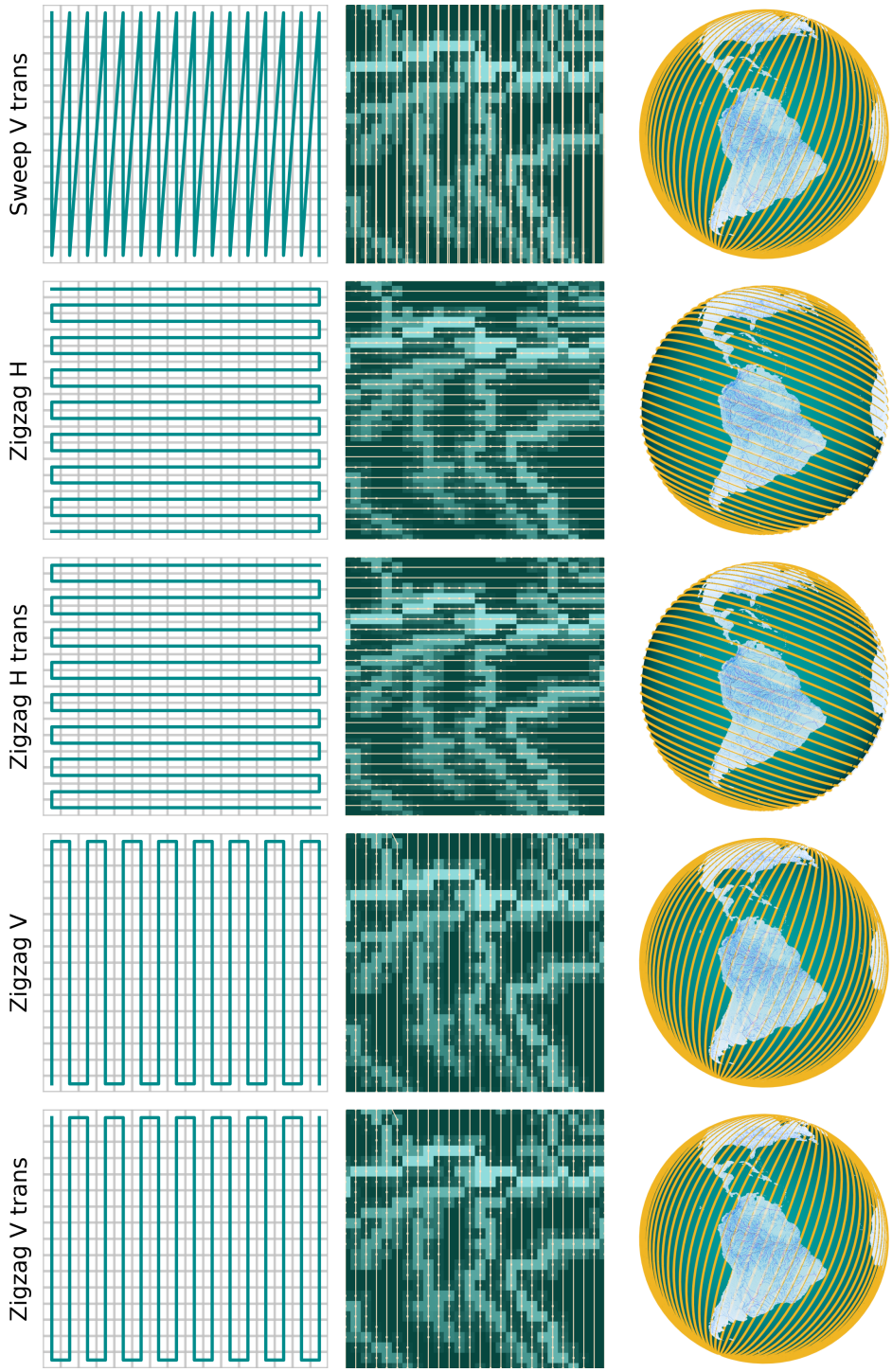


Figure 7.6: Visualization of different types of space-filling curves. For each type, we show the space-filling curve over a 2D discrete space (left), zoomed in version over the Earth where the points are sorted via a specific serialization order within the space-filling curve (middle), and simplified 3D visualization of the curve over the Earth (right).

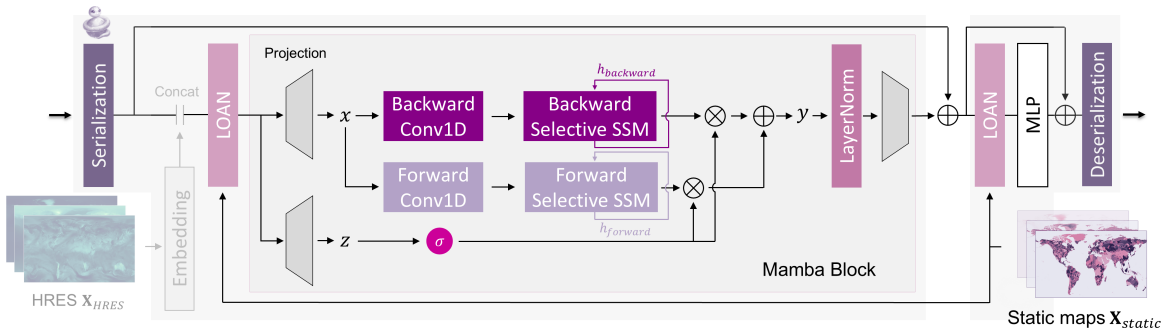


Figure 7.7: The structure of the hindcast block and forecast block. Both use a bidirectional Mamba block and the forecast block has the same structure as the hindcast block, but it additionally incorporates meteorological forecasts (HRES) by concatenation. The forecast block also includes LOAN layers although it is not shown in Fig. 7.2.

features are normalized by the LOAN layer, the Mamba block processes the features of the sampled points along the sequence (see Section 2.1.7).

Fig. 7.7 illustrates the steps of the Mamba block. The normalized sequence $\mathbf{X} \in \mathbb{R}^{N\times(T\times P)\times K}$ is projected into $\mathbf{x} \in \mathbb{R}^{N\times(T\times P)\times K}$ and $\mathbf{z} \in \mathbb{R}^{N\times(T\times P)\times K}$, where $T\times P$ is the length of the sequence. Note that the order of the elements in the sequence depends on the serialization, which differs between the hindcast blocks. We use a bi-directional approach that converts $\mathbf{x}$ into $\mathbf{x}'_o$ using a forward and a backward 1-D causal convolution, where $o \in \{f, b\}$ denotes the forward or backward pass. For each direction, $\mathbf{B}_o$, $\mathbf{C}_o$, and Δ_o are obtained by projection layers from $\mathbf{x}'_o$, and $\bar{\mathbf{A}}_o$ and $\bar{\mathbf{B}}_o$ are computed using the SSM equation (Eq. (2.19) in Section 2.1.7):

$$\bar{\mathbf{A}} = e^{(\Delta\mathbf{A})}, \quad \bar{\mathbf{B}} = (\Delta\mathbf{A})^{-1}(e^{(\Delta\mathbf{A})} - \mathbf{I})\Delta\mathbf{B}. \tag{7.6}$$

The selective SSM then uses the following equation (Eq. (2.18) in Section 2.1.7) to obtain $\mathbf{y}_{forward}$ and $\mathbf{y}_{backward}$ for the forward and backward pass, respectively:

$$h_t = \bar{\mathbf{A}}h_{t-1} + \bar{\mathbf{B}}x_t, \quad y_t = \mathbf{C}h_t + \mathbf{D}x_t. \tag{7.7}$$

The final output $\mathbf{y}$ is obtained by gating $\mathbf{y}_{forward}$ and $\mathbf{y}_{backward}$ via SiLU($\mathbf{z}$) and adding them up. Finally, $\mathbf{y}$ is normalized and projected back linearly to $\mathbb{R}^{N\times(T\times P)\times K}$. The complete algorithm for the Mamba block is described in Algorithm 1.

After the Mamba block, the hindcast block includes another LOAN layer followed by an MLP. The final output $\mathbf{X}$ is then deserialized at the end since the next hindcast block uses a different serialization.

7.2.6 Forecasting Layer

While the hindcast layers encode the sequence of past input variables into a K-dimensional vector per sampled point, e.g., $\mathbf{X}_{hindcast} \in \mathbb{R}^{N\times1\times P\times K}$, the forecasting layers forecast the difference of daily mean river discharge $\Delta\mathbf{X}^{t+l}_{dis24}$ for each lead time l, using $\mathbf{X}_{hindcast}$ and meteorological forecasts $\mathbf{X}^{t+1:t+L}_{HRES}$ as input, as shown in Fig. 7.2. The forecast blocks have the same structure as the hindcast blocks except that the forecast block incorporates the meteorological forcing (HRES). This is done by projecting $\mathbf{X}^{t+l}_{HRES}$ with a linear layer to 64 dimensions, serializing it, and concatenating it with

Algorithm 1 Mamba block

Require:
token sequence $\mathbf{X}^l : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{K})$
token sequence $\mathbf{X}_{static} : (\mathtt{N}, \mathtt{P}, \mathtt{K})$
curve ID (S_{block}) specific to the block l

Ensure:
transformed token sequence $\mathbf{X}^{l+1} : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{K})$

---

serialize the input sequence along the P dimension
$\mathbf{X}^l : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{K}), \mathbf{X}_{static} : (\mathtt{N}, \mathtt{P}, \mathtt{V_s}) \leftarrow \mathbf{Serialization}(\mathbf{X}^l, S_{block}), \mathbf{Serialization}(\mathbf{X}_{static}, S_{block})$
adaptively normalize the input sequence $\mathbf{X}^l$
$\mathbf{X}^{l'} : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{K}) \leftarrow \mathbf{LOAN}_1(\mathbf{X}^l, \mathbf{X}_{static})$
projection of $\mathbf{X}^{l'}$ into $\mathbf{x}$ and $\mathbf{z}$, here E is equal to K in our work since we do not expand the dimension
$\mathbf{x} : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{E}), \mathbf{z} : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{E}) \leftarrow \mathbf{Linear}^{\mathbf{xz}}(\mathbf{X}^{l'})$
# process with different direction
for o in {forward, backward} **do**
  # flip the curve along the spatial dimension P
  if d = 'backward' **then**
    $\mathbf{x} : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{E}) \leftarrow \mathbf{Flip}(\mathbf{x})$
  end if
  # flatten the curve along the temporal dimension 'spatial-first'
  $\mathbf{x}' : (\mathtt{N}, (\mathtt{T} \times \mathtt{P}), \mathtt{E}) \leftarrow \mathbf{Flatten}(\mathbf{x})$
  # selective state space model, here M is the D_state
  $\mathbf{x}'_o : (\mathtt{N}, (\mathtt{T} \times \mathtt{P}), \mathtt{E}) \leftarrow \mathbf{SiLU}(\mathbf{Conv1d}_o(\mathbf{x}'))$
  $\mathbf{B}_o : (\mathtt{N}, (\mathtt{T} \times \mathtt{P}), \mathtt{M}), \mathbf{C}_o : (\mathtt{N}, (\mathtt{T} \times \mathtt{P}), \mathtt{M}) \leftarrow \mathbf{Linear}^{\mathbf{B}}_o(\mathbf{x}'_o), \mathbf{Linear}^{\mathbf{C}}_o(\mathbf{x}'_o)$
  # initialize D_o with ones
  $\mathbf{D}_o : (\mathtt{E}) \leftarrow \mathbf{Parameter}$ Ones : $(\mathtt{E})$
  # softplus ensures positive $\mathbf{\Delta}_o$
  $\mathbf{\Delta}_o : (\mathtt{N}, (\mathtt{T} \times \mathtt{P}), \mathtt{E}) \leftarrow \log(1 + \exp(\mathbf{Linear}^{\mathbf{\Delta}}_o(\mathbf{x}'_o) + \mathbf{Parameter}^{\mathbf{\Delta}}_o))$
  # shape of $\mathbf{Parameter}^{\mathbf{A}}_o$ is $(\mathtt{E}, \mathtt{M})$
  $\overline{\mathbf{A}_o} : (\mathtt{N}, (\mathtt{T} \times \mathtt{P}), \mathtt{E}, \mathtt{N}) \leftarrow \mathbf{\Delta}_o \bigotimes \mathbf{Parameter}^{\mathbf{A}}_o$
  $\overline{\mathbf{B}_o} : (\mathtt{N}, (\mathtt{T} \times \mathtt{P}), \mathtt{E}, \mathtt{M}) \leftarrow \mathbf{\Delta}_o \bigotimes \mathbf{B}_o$
  # initialize h_o and $\mathbf{y}_o$ with zeros
  $h_o : (\mathtt{N}, \mathtt{E}, \mathtt{M}) \leftarrow$ Zeros : $(\mathtt{N}, \mathtt{E}, \mathtt{M})$
  $\mathbf{y}_o : (\mathtt{N}, (\mathtt{T} \times \mathtt{P}), \mathtt{E}) \leftarrow$ Zeros : $(\mathtt{N}, (\mathtt{T} \times \mathtt{P}), \mathtt{E})$
  # SSM recurrent
  for i in {0, ..., L-1} **do**
    $h_o = \overline{\mathbf{A}_o}[:, i, :, :] \odot h_o + \overline{\mathbf{B}_o}[:, i, :, :] \odot \mathbf{x}'_o[:, i, :, \mathtt{None}]$
    $\mathbf{y}_o[:, i, :] = h_o \bigotimes \mathbf{C}_o[:, i, :] + \mathbf{D}_o[\mathtt{None}, :] \odot \mathbf{x}'_o[:, i, :]$
  end for
  # reshape (T × P) to (T, P)
  $\mathbf{y_o} : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{E}) \leftarrow \mathbf{Reshape}(\mathbf{y_o})$
  # flip the curve along the spatial dimension P
  if o = 'backward' **then**
    $\mathbf{y_o} : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{E}) \leftarrow \mathbf{Flip}(\mathbf{y_o})$
  end if
end for
# get gated $\mathbf{y}$
$\mathbf{y}'_{forward} : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{E}), \mathbf{y}'_{backward} : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{E}) \leftarrow \mathbf{y}_{forward} \odot \mathbf{SiLU}(\mathbf{z}), \mathbf{y}_{backward} \odot \mathbf{SiLU}(\mathbf{z})$
post normalization and residual connection
$\mathbf{X}^{l+1'} : (\mathtt{N}, (\mathtt{T} \times \mathtt{P}), \mathtt{K}) \leftarrow \mathbf{Linear}^{\mathbf{X}}(\mathbf{LayerNorm}((\mathbf{y}'_{forward} + \mathbf{y}'_{backward})/2)) + \mathbf{X}^l$
adaptively normalize the output sequence $\mathbf{X}^{l+1}$
$\mathbf{X}^{l+1} : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{K}) \leftarrow \mathbf{LOAN}_2(\mathbf{X}^{l+1'}, \mathbf{X}_{static})$
feed-forward layer and residual connection
$\mathbf{X}^{l+1} : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{K}) \leftarrow \mathbf{MLP}(\mathbf{X}^{l+1}) + \mathbf{X}^{l+1'}$
resort the input sequence along the P dimension
$\mathbf{X}^{l+1} : (\mathtt{N}, \mathtt{T}, \mathtt{P}, \mathtt{K}) \leftarrow \mathbf{Resort}(\mathbf{X}^{l+1}, S_{block})$
Return: $\mathbf{X}^{l+1}$

the input $\mathbf{X}$ as illustrated in Fig. 7.7. The processing of HRES is done sequentially, e.g., we have L forecast blocks and the l-th forecast block processes $\mathbf{X}^{t+l}_{HRES}$. We argue that this design is crucial to ensure that the temporal relationships between the meteorological forcing and the initial conditions are maintained.

The output of all forecast blocks is processed by L regression heads implemented as multi-layer perceptrons (MLP) where the output for the lead time $t + l$ is obtained as:

$$\Delta\mathbf{X}^{t+l}_{dis24} = \text{Linear}\Big(\text{ReLU}\Big(\text{Concat}\big(\text{Linear}(\mathbf{X}^{t+l}_{forecast}), \text{Linear}(\mathbf{X}^{t+1:t+L\setminus t+l}_{forecast})\big)\Big)\Big)\,, \tag{7.8}$$

where $\mathbf{X}^{t+l}_{forecast}$ are the features from the l-th forecast block and $\mathbf{X}^{t+1:t+L\setminus t+l}_{forecast}$ are the concatenated features from all forecast blocks except of the l-th block. The linear layers project the input $\mathbf{X}^{t+l}_{forecast}$ or $\mathbf{X}^{t+1:t+L\setminus t+l}_{forecast}$ to 32 dimensions and the last linear projection estimates finally $\Delta\mathbf{X}^{t+l}_{dis24} \in \mathbb{R}^{N\times 1\times P\times 1}$.

7.2.7 Training

As already mentioned, we sample P points around the globe for training. As a target value for training, we first use the river discharge data from the GloFAS reanalysis as ground truth and then fine-tune on sparse observations using data from the Global Runoff Data Centre (GRDC, Section 2.2.7). For GRDC fine-tuning, we take P as the number of input points per sample and compute the loss only on points where GRDC observations are available without considering reanalysis data from GloFAS. We obtain the target values by $\Delta\hat{\mathbf{X}}^{t+l}_{dis24}(p) = \hat{\mathbf{X}}^{t+l}_{dis24}(p) - \hat{\mathbf{X}}^{t-1}_{dis24}(p)$, where $\hat{\mathbf{X}}$ are the values from GloFAS or GRDC. For the training loss, we propose a weighted version of the mean-squared error (MSE) loss:

$$\mathcal{L} = \frac{1}{N\times P\times L}\sum_{n=1}^{N}\sum_{p=1}^{P}\sum_{l=1}^{L} w^{n,t+l}(p)\|\Delta\hat{\mathbf{X}}^{n,t+l}_{dis24}(p) - \Delta\mathbf{X}^{n,t+l}_{dis24}(p)\|_2^2\,, \tag{7.9}$$

where N is the batch size. Since the severity of a flood is highly important for flood forecasting and severe floods occur rarely, the weighting factor $w^{n,t+l}(p)$ takes this into account. The severity of a flood is ranked by the statistical flood return period in years, which we denote by r and ranges from 1.5 to 500. These ranges are also used in GloFAS. We note that a high return period event simply reflects statistical rarity in streamflow magnitude, and should not be equated with a flood event without additional context, e.g., thresholds or inundation. The return period is used here as a proxy indicator of hydrological extremity, which we call flood. The severity of a flood is thus given by $\hat{r}^{t+l}(p) = \max_r \left\{r : \hat{\mathbf{X}}^{t+l}_{dis24}(p) \geqslant \theta_r(p)\right\}$, where θ_r is the statistical threshold for a given flood return period r. We also include the case $r{=}0$ with $\theta_r{=}0$ for defining events that are not floods. Using this notation, the weighting is thus given by

$$\hat{w}^{n,t+l}(p) = \begin{cases} \hat{r}^{n,t+l}(p) & \text{if} \quad \hat{r}^{n,t+l}(p) > 1 \\ 1 & \text{otherwise.} \end{cases} \tag{7.10}$$

We thus weight the loss based on the flood return period if a flood occurred at location p and time $t + l$, and we use 1 if there has been no flood. We further weight the loss with $\hat{u}^{b,t+l} = e^{\alpha(L-l+1)}$, where we give a higher weight to a shorter lead time l and use $\alpha{=}0.25$. This compensates for the

sequential structure of the forecast blocks where each forecast block takes the features of the previous block as input. The final weight is thus given by $w^{n,t+l}(p) = \hat{u}^{n,t+l}\hat{w}^{n,t+l}(p)$. Since river discharge exhibits a very large dynamic with varying orders of magnitude, we transform the discharge values by $\text{sign}(\Delta\hat{x})\log(1 + |\Delta\hat{x}|)$. We evaluate the impact of the weighting and provide more details in Section 7.3.8.

For inference, we can forecast floods for any set of points or densely as in Fig. 7.1. For training and inference on the global dataset, it is impractical to fit all the input points ($\sim$ 6 million points) into the memory. For this, we first define a Gilbert space-filling curve on the globe and then we split the curve into smaller curves along the space-filling curve, e.g., we split the curve into sequences with $\sim$ 311K points for each. A simplified version of the splits is shown in Fig. 7.8.

Figure 7.8: A simplified view of splitting along Gilbert space-filling curve.

7.3 Experimental results

This section is structured as follows. We describe the experimental setting in Section 7.3.1. Then we evaluate the model in gauged setting on GloFAS reanalysis data in Section 7.3.2 and on GRDC observational data in Section 7.3.3. In Section 7.3.4, we compare to the Google reforecast in ungauged setting and in Section 7.3.5, we compare to the Google reforecast in gauged setting. In Section 7.3.6, we evaluate on the operational forecast of GloFAS. Section 7.3.7 shows case studies of extreme floods. Finally, Section 7.3.8 includes the ablation studies and analyses.

7.3.1 Experimental Setup

7.3.1.1 Datasets

We obtain data for river discharge from the ECMWF GloFAS reanalysis (*Harrigan et al.*, 2020) (Section 2.2.6). It is generated by forcing the LISFLOOD hydrological model (*Knijff et al.*, 2010) using meteorological data from ERA5 (*Hersbach et al.*, 2020). GloFAS reanalysis combines physics-based simulation with observations to generate a consistent reconstruction of the past. The dataset is provided as a daily averaged discharge on a global coverage at 3 arcmin grid (0.05°). We use the

GloFAS reanalysis as a target discharge for training and testing the model. In addition, we fine-tune and test the model on observational GRDC river discharge data (Section 2.2.7). The initial conditions for the forecast consist of data from the GloFAS reanalysis, ERA5-Land reanalysis (Section 2.2.4.2), and observational precipitation estimates from NOAA CPC (Section 2.2.9). The meteorological forcing for the forecasting is based on the medium-range weather forecasts from ECMWF IFS and HRES atmospheric model (Section 2.2.8). The static features are derived from LISFLOOD (Section 2.2.10). We also use HydroRIVERS static data (Section 2.2.11) for a comparison in the ablation studies.

Flood thresholds are determined using return periods for individual points and are calculated from the long-term data (Section 2.3). The thresholds allow for the identification of a flood when the threshold is surpassed. In order to run experiments efficiently, we sampled points as showed in Section 2.2.12. Note that the trained model can generate river discharge maps at full resolution as can be seen in Sec. 7.3.7.

7.3.1.2 Implementation Details

The training was done on clusters with NVIDIA A100 80GB and 48GB GPUs. In Table 7.1, we highlight the main hyperparameters used for training RiverMamba.

To mimic a real operational setting, the initial conditions from CPC data starts at day $t-2$, GLoFAS reanalysis at day $t-1$, and ERA5-Land reanalysis at day $t-1$ in the past. All input data are normalized based on the computed mean and standard deviation from the training set. To handle missing data in the reanalysis, we first use the pre-computed statistics to normalize the data. Then, we replace the invalid pixels with zero values. HRES data is always used for validating and testing. During training, we replace IFS meteorological forcing by ERA5 if they are unavailable, e.g., before 2010. We train on the years 1979-2018, validate on 2019-2020, and test on 2021-2024. The training, validation, and testing splits are shown in Fig. 7.9.

To accelerate training and to fit the data into the memory, we use bfloat16 floating point precision. During inference, we use float32 floating point precision. Pre-training RiverMamba took about 3 days on 16 GPUs. Finetuning on GRDC data took about 4 hours on 16 GPUs. Additional details and code are available at `https://github.com/HakamShams/RiverMamba_code`.

7.3.1.3 Baselines

We compare RiverMamba to persistence, climatology, and the state-of-the-art deep learning Encoder-Decoder LSTM of Google's operational flood forecasting system (*Nearing et al.*, 2024). For evaluation on GRDC observations, we additionally compare our approach to the reforecast version of the state-of-the-art operational GloFAS forecasting system operated by ECMWF (*Harrigan et al.*, 2023).

Climatology. We followed *Harrigan et al.* (2023) to define the climatology baseline. For this, we computed climatology for the long-term record of river discharge data (1979-2018) with a moving window of 31 days centered on the day-of-the-year. Then, we computed 11 fixed quantiles at 10% interval for each day-of-the-year. As a result, the climate distribution changes with lead time, reflecting the dynamic changes in local river discharge patterns over time. Climatology is commonly applied to medium- and extended range lead times, where seasonal patterns predominantly influence the river discharge forecast (*Harrigan et al.*, 2023).

Configuration	Pre-training (GloFAS-Reanalysis)	Fine-tuning (GRDC)
Optimizer	AdamW	AdamW
Learning rate	0.0006	0.0001
Minimum learning rate	0.00009	0.00009
Batch size (B)	1	1
Learning rate scheduler	Cosine annealing	Cosine annealing
Weight decay	0.001	0.01
Training epochs	60	20
Warmup epochs	4	6
Gradient clip	10	10
Input hindcast length (T)	4	4
Lead time (L)	7	7
α for $\hat{u}$	0.25	0.25
Number of input points P	245,954	245,954*
Embedding dimension for GloFAS reanalysis	48	48
Embedding dimension for ERA5-Land reanalysis	128	128
Embedding dimension for CPC	16	16
Number of hindcast layers	3	3
Hidden dimension in hindcast block (K)	192	192
Depth of hindcast layers	[2, 2, 2]	[2, 2, 2]
Curves in hindcast layers	{Sweep_H, Sweep_V, Gilbert, Gilbert trans}	{Sweep_H, Sweep_V, Gilbert, Gilbert trans}
Grouping size in hindcast block	[(4, 254945), (2, 254945), (1, 254945)]	[(4, 254945), (2, 254945), (1, 254945)]
Dropout in hindcast block	0.2	0.4
D_state in hindcast block	16	16
D_conv in hindcast block	4	4
Hidden dimension in forecast block (K)	192+64	192+64
Embedding dimension for HRES (K_{HRES})	64	64
Number of forecast layers	7	7
Depth of forecast layers	[1, 1, 1, 1, 1, 1, 1]	[1, 1, 1, 1, 1, 1, 1]
Curves in forecast layers	{Sweep_H, Sweep_V, Gilbert, Gilbert trans}	{Sweep_H, Sweep_V, Gilbert, Gilbert trans}
Grouping size in forecast block	[(1, 254945)] * 7	[(1, 254945)] * 7
Dropout in forecast block	0.2	0.4
D_state in forecast block	16	16
D_conv in forecast block	4	4
Hidden dimension in forecasting head (K_{head})	64	64
Dropout in head	0.1	0.3

*For GRDC fine-tuning, we only compute the loss where GRDC observations are available.

Table 7.1: Implementation details of RiverMamba

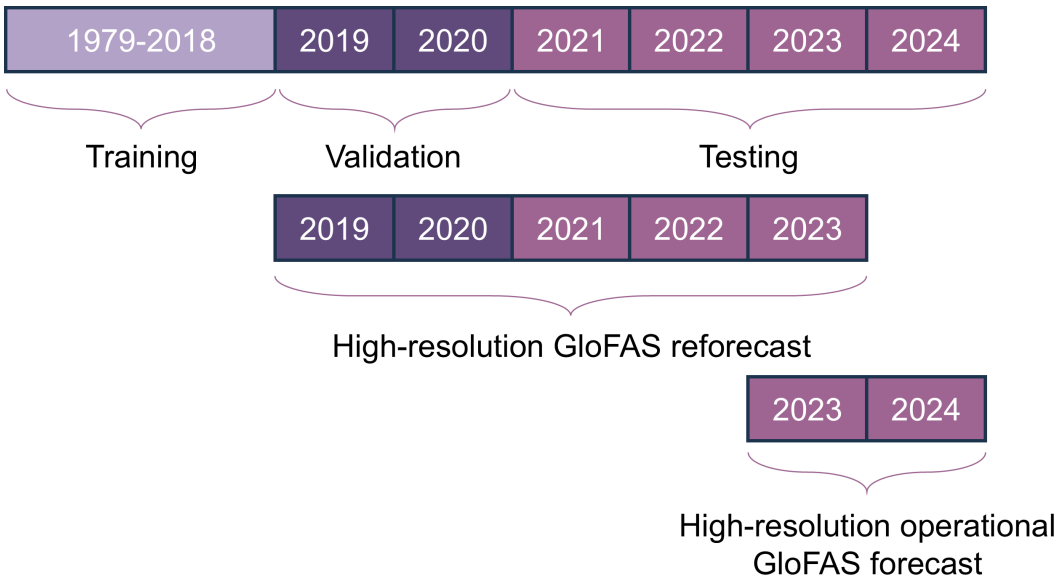


Figure 7.9: Details about the data splits.

Persistence. We defined the persistence baseline as the daily river discharge of GloFAS reanalysis from the day preceding the day at which the forecast was issued, e.g., for a forecast starting at 00:00 UTC $\mathbf{X}^t$, the persistence is defined as the averaged river discharge between 00:00 UTC $\mathbf{X}^{t-1}$ and 00:00 UTC $\mathbf{X}^t$. This value was used as a prediction for the entire lead time. Persistence is primarily applied to short lead times, where the correlation of sequential river discharge values predominantly influences the forecasts (*Harrigan et al.*, 2023). Note that this baseline is unrealistic since no reanalysis is available directly at time t.

LSTM. We adopted the same LSTM architecture as described in *Nearing et al.* (2024). For the LSTM model, we followed the same protocol as originally proposed in *Nearing et al.* (2024), which considers only temporal context but does not include any spatial connections. The space filling curves are thus not used in combination with the LSTM baseline. To ensure a consistent evaluation, we trained LSTM on the same input data as RiverMamba. All results in Sections 7.3.2, 7.3.3, and 7.3.6 are obtained with our trained LSTM. A comparison with the published reforecasts by Google's LSTM (*Nearing et al.*, 2024) is available in Sections 7.3.4 and 7.3.5.

The model follows an encoder-decoder structure, where the encoder is a bi-directional "hindcast" LSTM that processes historical input data, and the decoder is a uni-directional "forecast" LSTM that generates predictions over a 7-day forecast horizon based on forecast inputs. To ensure fair comparison and benchmarking, we used the same input data (e.g., we include GloFAS reanalysis and exclude IMERG and nowcasting data for LSTM), train-test split, and normalization strategies as in the RiverMamba model. To remain consistent with *Nearing et al.* (2024), we trained the model only at locations with available gauge observations-specifically the 3366 GRDC stations rather than using a global training setup. Thus for LSTM, we did not include any spatial connections and space filling curves are not used in combination with the LSTM baseline.

The model leverages both dynamic and static inputs. For the hindcast LSTM, we used a 14-day sequence of dynamic inputs including CPC precipitation, GloFAS reanalysis, and ERA5-Land

reanalysis data. At each time step, static attributes derived from the LISFLOOD model are embedded and concatenated with the dynamic inputs. For the forecast LSTM, we used ECMWF HRES forecasts as dynamic inputs over the 7-day horizon, with the static attributes concatenated in the same manner.

To connect the encoder and decoder, we employed a "state" layer consisting of two transfer networks (https://neuralhydrology.github.io/): a linear cell-state transfer network and a nonlinear hidden-state transfer network (a fully connected layer with hyperbolic tangent activation). A linear output head is applied at each forecast step to predict streamflow, and the model is trained using the mean squared error (MSE) loss. Unlike *Nearing et al.* (2024), we focus on deterministic prediction, so we do not implement a probabilistic output head or probabilistic loss function. In total, the model has $834,421$ parameters.

In *Nearing et al.* (2024), an input sequence length of 365 days was used. This is because the model in *Nearing et al.* (2024) has to simulate the states (e.g., soil moisture) and current runoff from the meteorological forcing input. In our experiments, since the states and the streamflow already integrate the meteorological signal of the past, we trained the LSTM model using a range of input sequence lengths from 4 to 90 days. We observed only marginal performance gains beyond a certain point, and identified 14 days as an optimal input sequence length.

The reported LSTM results are averaged over an ensemble of three independently trained models, each initialized with a different random seed. Each training batch contains data from all 3366 GRDC stations at a given time step, with a batch size of 1, effectively training on 3366 samples per mini-batch. Training takes approximately 12 hours on four NVIDIA A100 GPUs for 35 epochs. More details about the model architecture can be found in *Nearing et al.* (2024), as well as in the NeuralHydrology GitHub repository (https://neuralhydrology.github.io/).

Table 7.2 summarizes the key hyperparameters used in our implementation of the LSTM model.

Configuration	Value
Hidden size in hindcast LSTM	256
Hidden size in forecast LSTM	128
Hidden size in static embedding layer	20
Hidden size in dynamic embedding layer	20
Hidden size in state layer	128
Number of layers	1
Dropout at output regression head	0.4
Dropout at state layer	0.1
Learning rate	0.0003
Learning rate scheduler	Cosine annealing
Batch size	1 with (3366 samples)
Optimizer	Adam
beta1 momentum term	0.9
beta2 momentum term	0.999
weight decay	0

Table 7.2: Implementation details of the LSTM model.

It is important to note that *Nearing et al.* (2024) did not release the full code or the full hyperparameter configurations of their final model, but only the pretrained checkpoints were made available. Although the saved models can be loaded for inference using the original inputs, it is not possible to

retrain or adapt these models to a different input setup, which was required for our experiments. We therefore used the published checkpoints and the NeuralHydrology GitHub repository as a reference to re-implement and train the LSTM.

GloFAS forecast. Operational forecast from GloFAS was obtained from the ECMWF Early Warning Data Store (EWDS) (https://doi.org/10.24381/cds.ff1aef77). This represents real-time data from the official system for operational flood forecasting from the Copernicus Emergency Management Service (CEMS) and managed and developed by the European Commission's Joint Research Centre. GloFAS forecast is produced by forcing the LISFLOOD model with the ECMWF ensemble forecast (ENS) up to 30 days. Note that GloFAS forecast uses ENS meteorological forcing daily at 00:00 UTC. The high-resolution GloFAS v.4.0 forecast is available from 2023-07-26. We compare to this baseline in Section 7.3.6.

GloFAS reforecast. This baseline is similar to GloFAS forecast, however, GloFAS reforecast are forecasts run over the past with the new system version 4.0. The reforecast is available until 2023 and does not span the full testing split. Reforecast from GloFAS was obtained from the ECMWF Early Warning Data Store (EWDS) (https://doi.org/10.24381/cds.2d78664e). We use this baseline for the main comparison with GloFAS in the main paper and in Section 7.3.3.

7.3.1.4 Evaluation Metrics

We evaluate the performance of RiverMamba on both GloFAS reanalysis and GRDC, where diagnostic GRDC stations are available (3366 stations). To assess model performance, we use metrics that are commonly used for hydrological modeling and flood forecasting evaluation (*Gauch et al.*, 2023). This includes MAE (Mean Absolute Error), RMSE (Root Mean Square Error), R (Pearson Correlation Coefficient), R2 (Coefficient of Determination), KGE (Kling-Gupta Efficiency), and the averaged Precision, Recall and F1-score for floods with return periods of 1.5 to 20 years. All evaluation points are gauged stations and temporally out-of-sample. The metrics are calculated on the time series at single grid points and then averaged over all points.

7.3.2 Experiments on Gauged GloFAS River Discharge Reanalysis

The quantitative results are shown in Table 7.3. As can be seen, the climatology baseline performs poorly, as the dynamic in local river discharge varies a lot over time, highlighting the difficulty in predicting flows. We therefore exclude it in Fig. 7.10 (a) that shows F1-score for floods with a 1.5-year return period and (b) KGE for river discharge for different lead times from 24 to 168 hours. The boxes show distribution quartiles and the evaluation points are represented as points along the y-axis.

Fig. 7.10 (d) shows the F1-score averaged over return periods of 1.5 to 20 years and (e) shows the median R2 for river discharge. The persistence baseline predicts the future discharge as the same value of the discharge at time t. This achieves good prediction for the short-term forecast, however, the prediction skill drops with lead time. While LSTM (*Nearing et al.*, 2024) outperforms the persistence baseline, RiverMamba outperforms all baselines and methods on all metrics as shown in Table 7.3. In particular for lead times above 48 hours, the performance gap between RiverMamba and LSTM is large. We attribute this to the receptive field and the spatio-temporal modeling of RiverMamba. Fig. 7.10 (c) plots the F1-score averaged over 24 to 168 hours lead time for different

	Validation (2019-2020)			Test (2021-2024)		
Model	R2(↑)	KGE(↑)	F1(↑)	R2(↑)	KGE(↑)	F1(↑)
Climatology	0.1175	0.2618	–	0.1352	0.2449	–
Persistence	0.6778	0.8380	0.3138	0.6833	0.8412	0.3223
LSTM	0.8539±0.0031	0.8931±0.0034	0.3511±0.0068	0.8485±0.0021	0.8924±0.0029	0.3582±0.0058
RiverMamba	**0.8803**±0.0043	**0.9137**±0.0026	**0.4540**±0.0056	**0.8728**±0.0013	**0.9125**±0.0008	**0.4589**±0.0080

Table 7.3: Results on GloFAS-Reanalysis. (±) denotes the standard deviation for 3 runs. F1-score (↑) is averaged for all lead times and 1.5-20 year return periods.

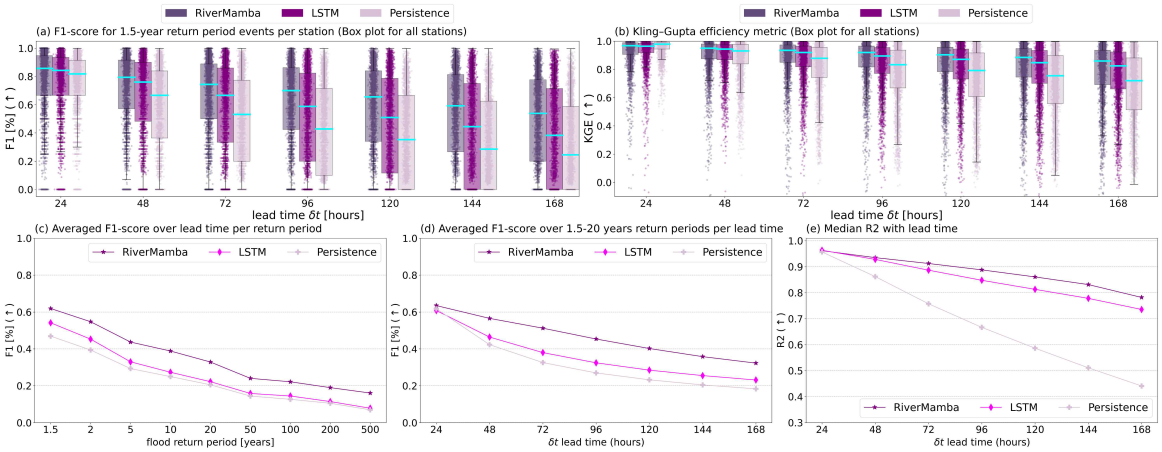


Figure 7.10: Results on GloFAS reanalysis across lead times and flood return periods.

flood return periods. The results show that RiverMamba outperforms the other approaches both for more frequent floods and rare severe floods that occur statistically only every 500 years.

In Figs. 7.11-7.23, we plot additional results for the experiments on GloFAS river discharge reanalysis. In Figs. 7.11-7.14, we report the results for MAE, RMSE, R2, and R metrics with lead time. As can be seen RiverMamba outperforms all baselines across these metrics. In Figs. 7.15-7.20, we report the results of F1-score, Precision, and Recall metrics for different return periods and lead times. From these figures, we can observe that RiverMamba improves the recall of floods in comparison to LSTM. However, LSTM has a better precision for the first lead time of 24 hours (Figs. 7.16 and 7.19). While RiverMamba has better F1 scores, especially, for longer lead time and with higher return periods (extreme floods). In Figs. 7.21 and 7.22, we compare the results between RiverMamba and LSTM for F1-score and KGE metrics spatially. The result shows that RiverMamba improves F1-score over LSTM in 41% of the stations and being better or equally better in 89% of the stations. It also shows that RiverMamba improves KGE over LSTM in 73% of the stations. Finally, Fig. 7.23 shows the confusion matrix for both RiverMamba and LSTM. From the comparison, we can see that RiverMamba improves the recall for extreme floods with higher return periods over LSTM, e.g., RiverMamba improves the true positive by 0.2 for floods with 500-year return period.

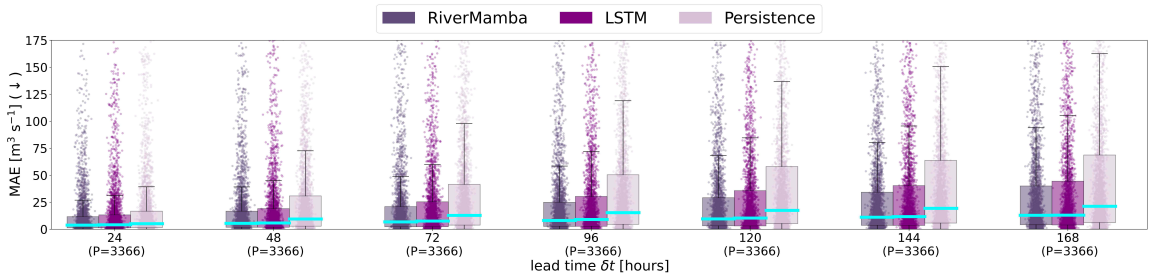


Figure 7.11: MAE of the river discharge forecasting with different lead time on GloFAS reanalysis (test set 2021-2024 temporally out-of-sample).

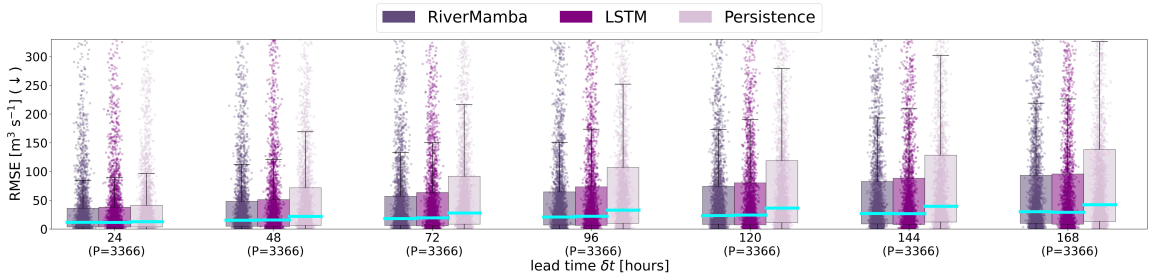


Figure 7.12: RMSE of the river discharge forecasting with different lead time on GloFAS reanalysis (test set 2021-2024 temporally out-of-sample).

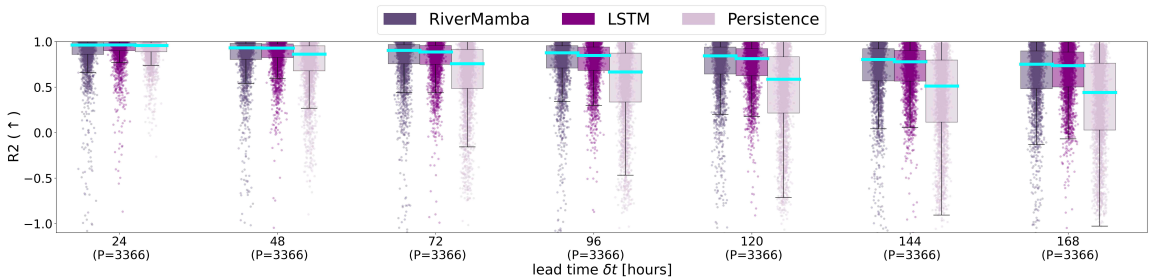


Figure 7.13: R2 (NSE) of the river discharge forecasting with different lead time on GloFAS reanalysis (test set 2021-2024 temporally out-of-sample).

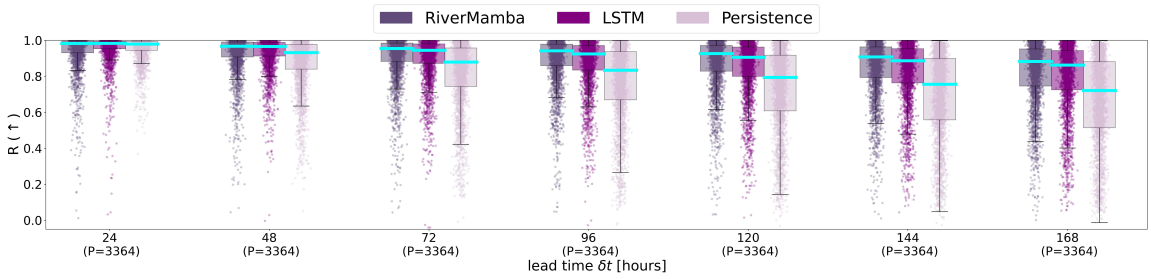


Figure 7.14: Pearson correlation (R) of the river discharge forecasting with different lead time on GloFAS reanalysis (test set 2021-2024 temporally out-of-sample).

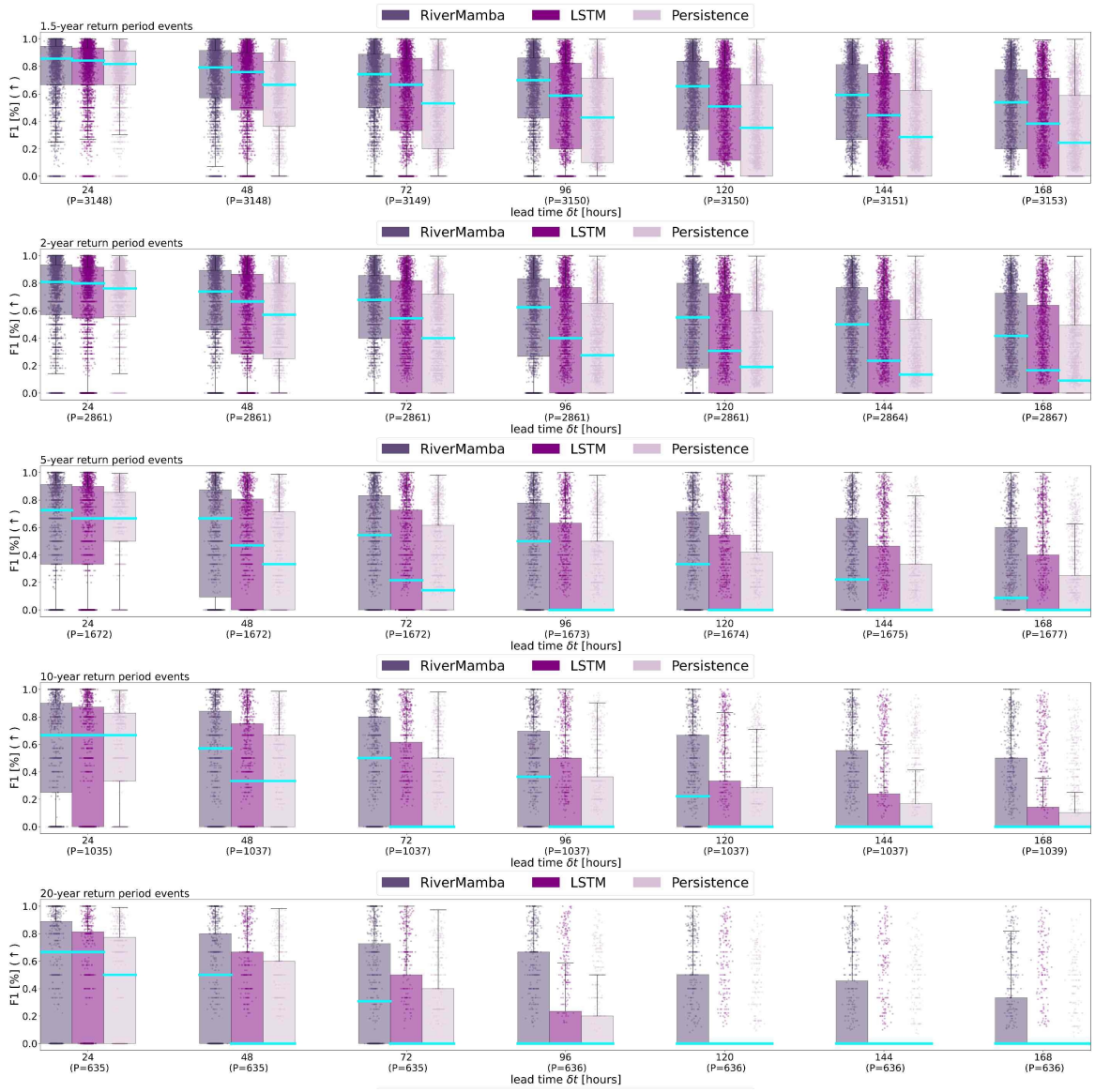


Figure 7.15: F1-score of flood forecasting for different return periods and lead time on GloFAS reanalysis (test set 2021-2024 temporally out-of-sample). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

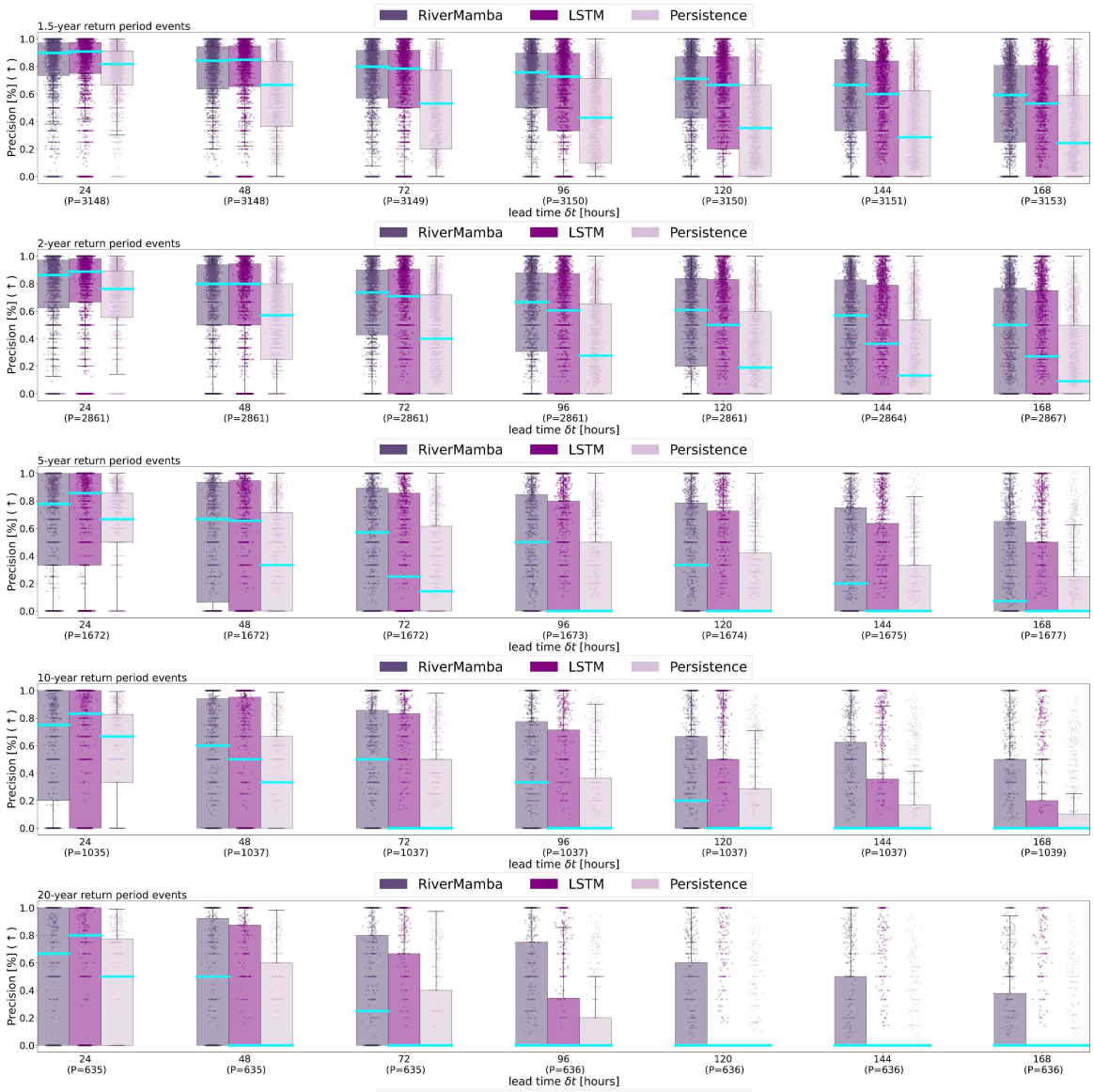


Figure 7.16: Precision of flood forecasting for different return periods and lead time on GloFAS reanalysis (test set 2021-2024 temporally out-of-sample). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

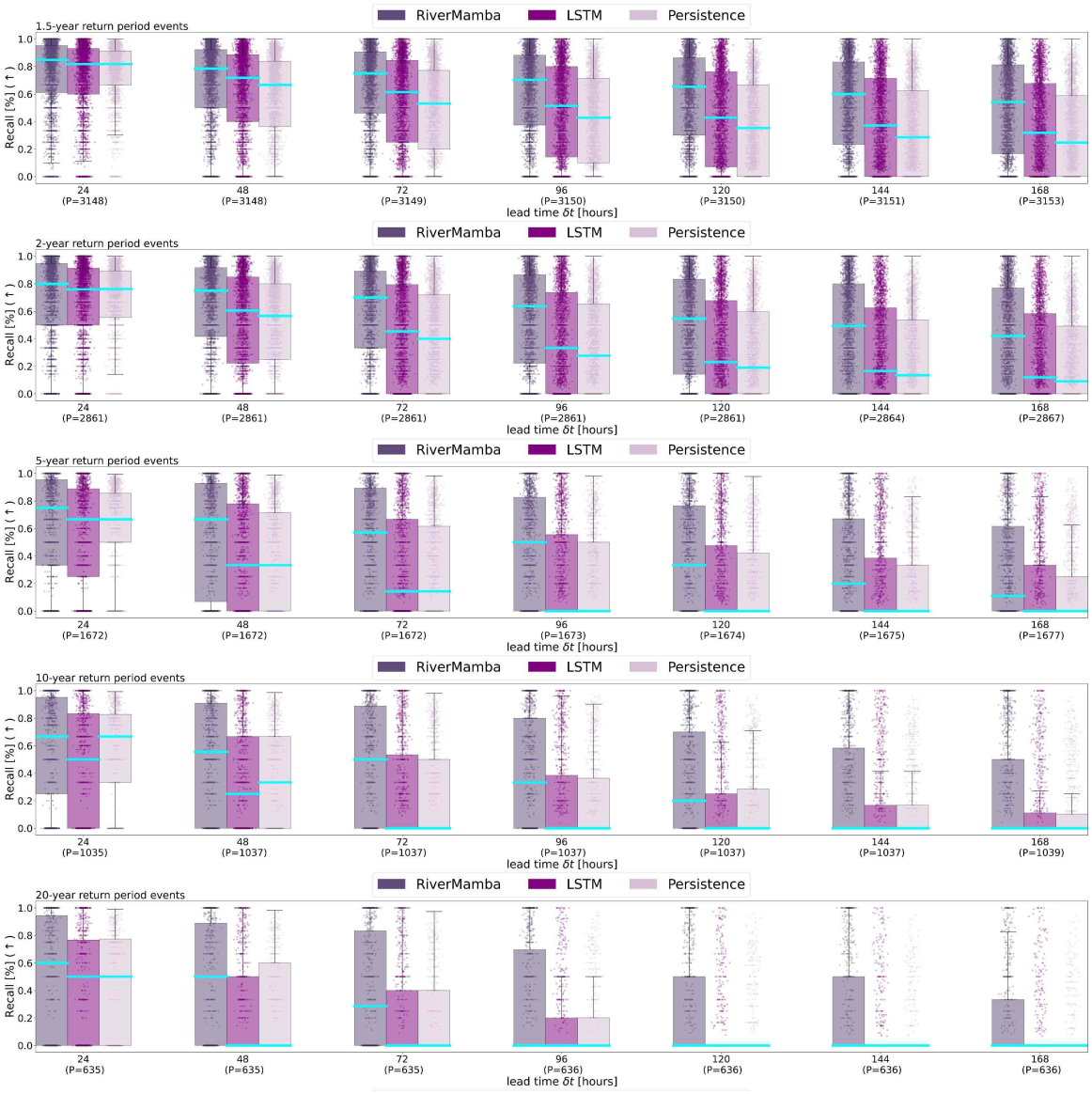


Figure 7.17: Recall of flood forecasting for different return periods and lead time on GloFAS reanalysis (test set 2021-2024 temporally out-of-sample). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

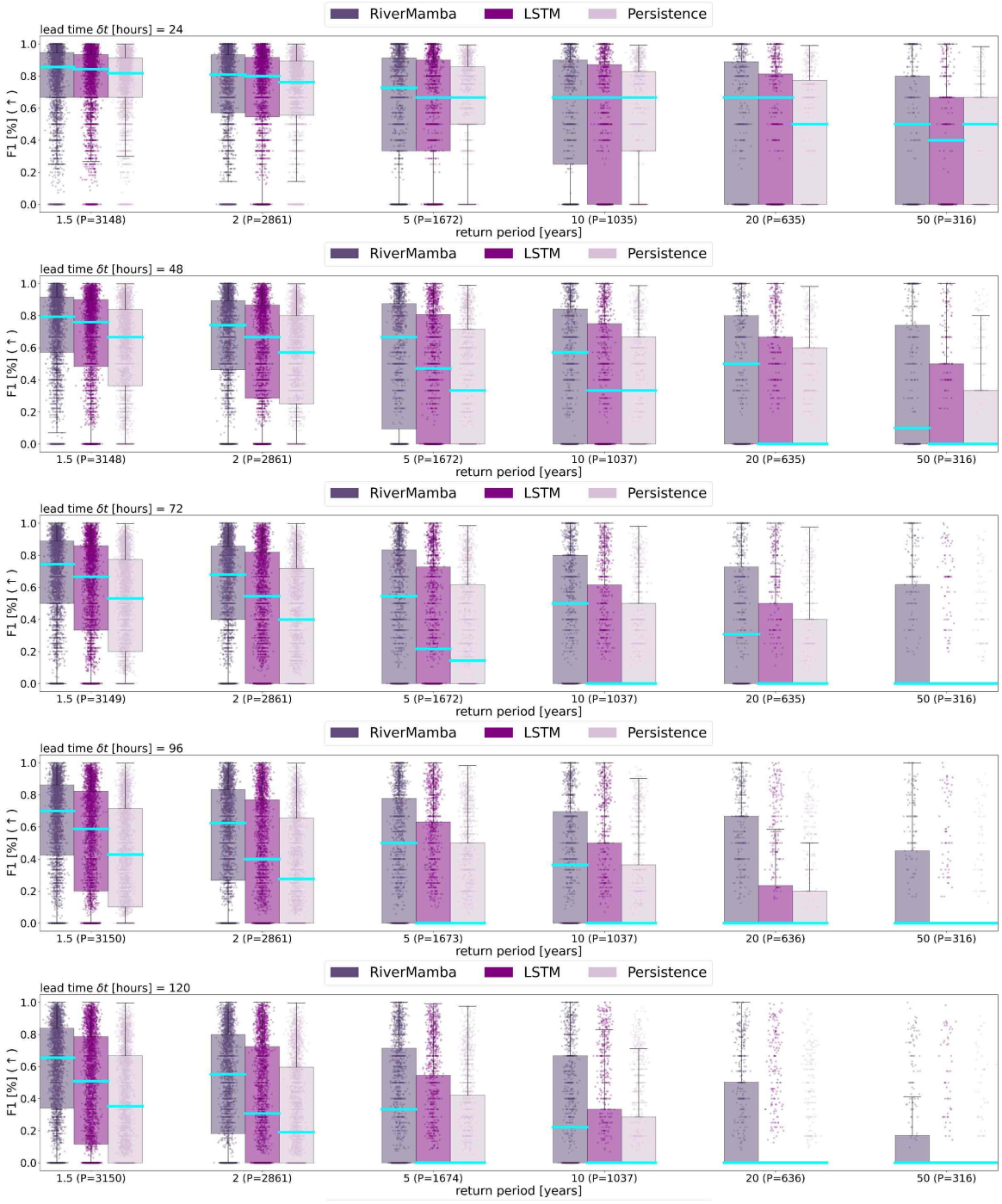


Figure 7.18: F1-score of flood forecasting for different lead time and return periods on GloFAS reanalysis (test set 2021-2024 temporally out-of-sample). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

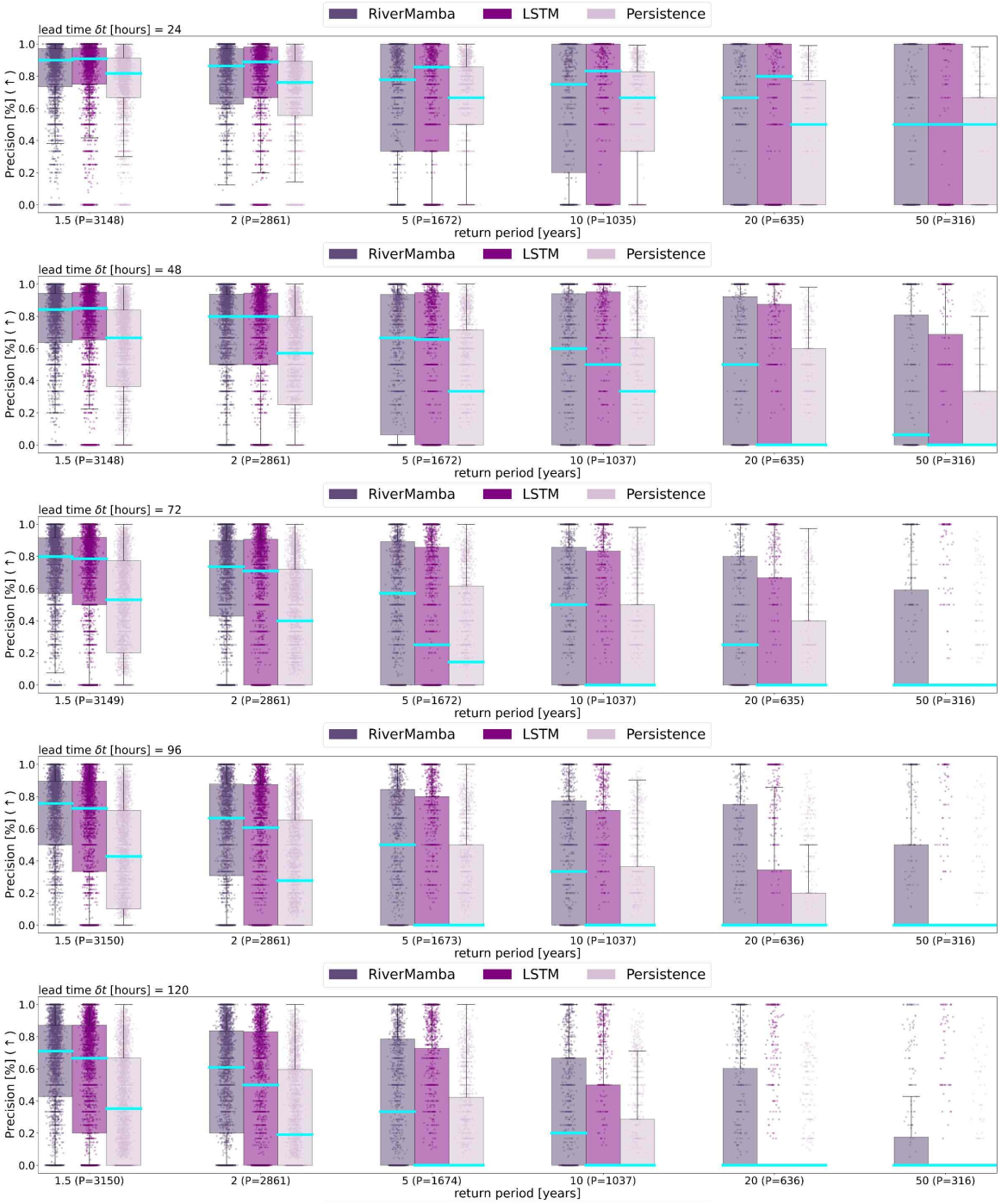


Figure 7.19: Precision of flood forecasting for different lead time and return periods on GloFAS reanalysis (test set 2021-2024 temporally out-of-sample). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

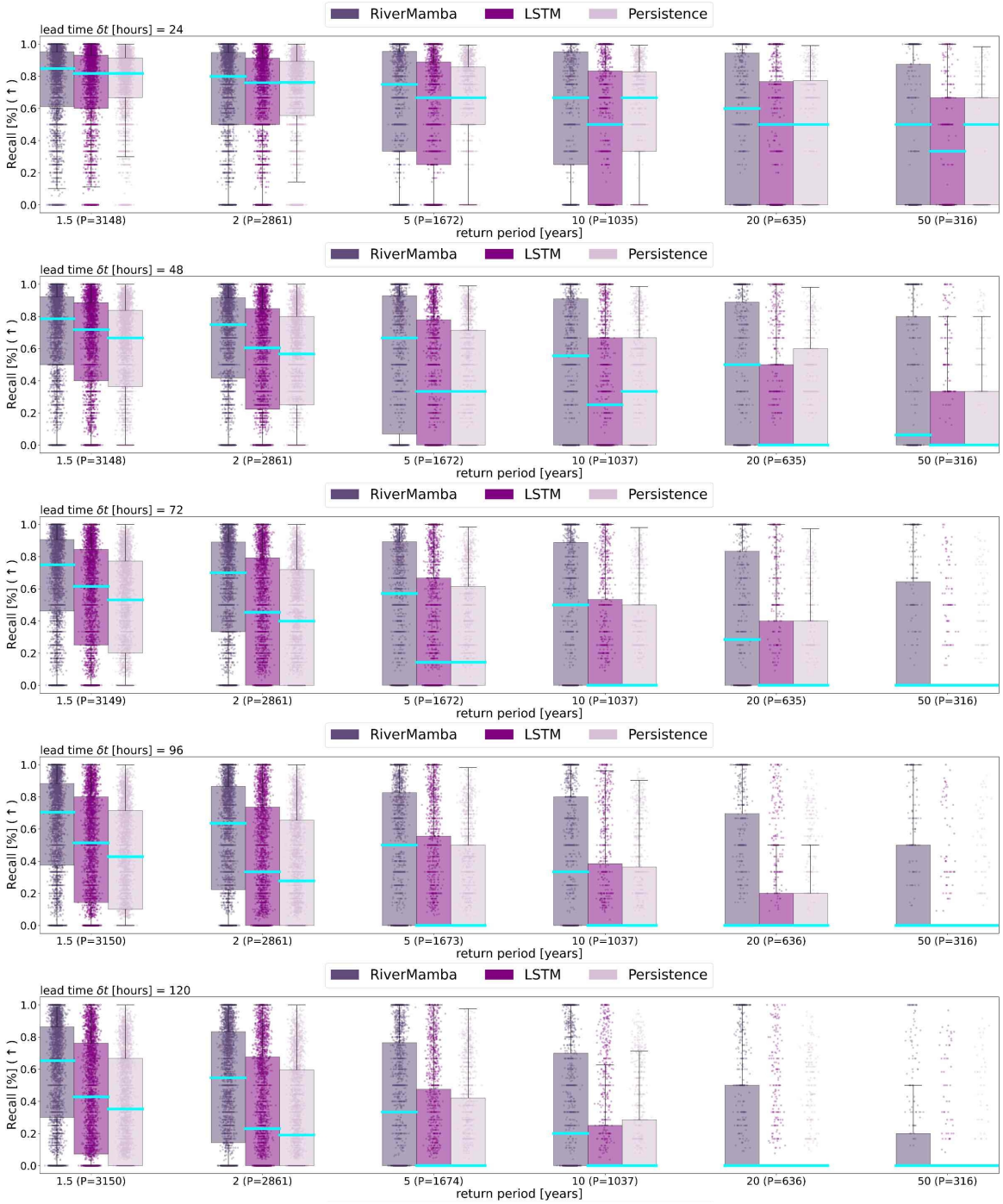


Figure 7.20: Recall of flood forecasting for different lead time and return periods on GloFAS reanalysis (test set 2021-2024 temporally out-of-sample). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

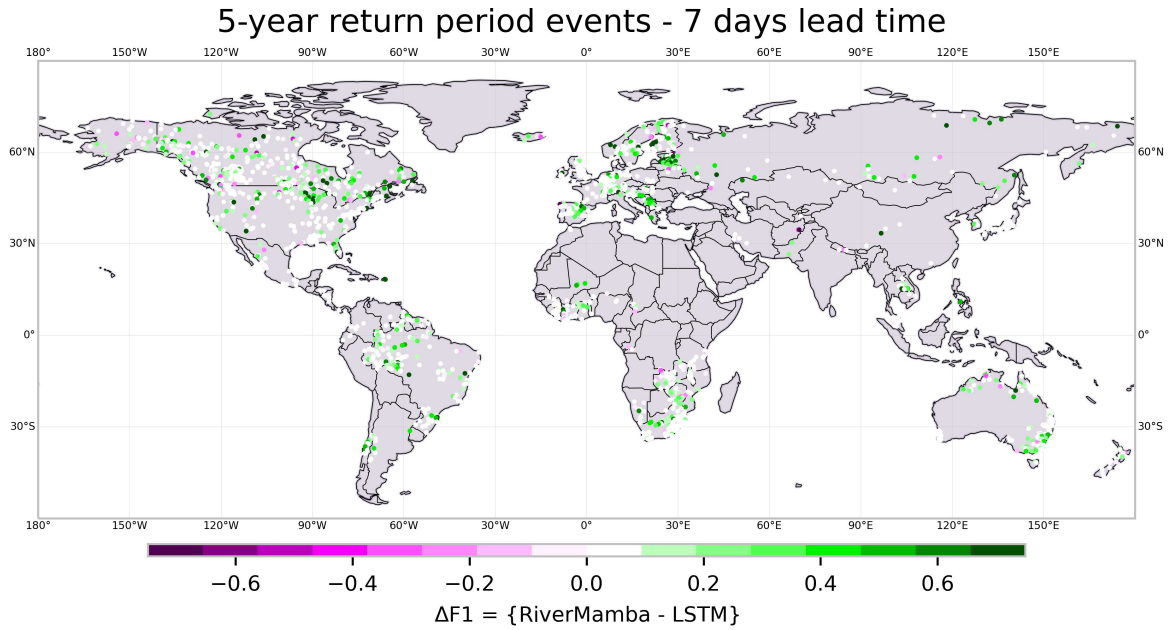


Figure 7.21: Comparison of F1-score between RiverMamba and LSTM (Google) on GloFAS reanalysis for the 5-year return period events (test set 2021-2024 temporally out-of-sample). RiverMamba improves over LSTM in 41% of the stations (P=1677) and being better or equally better in 89% of the stations.

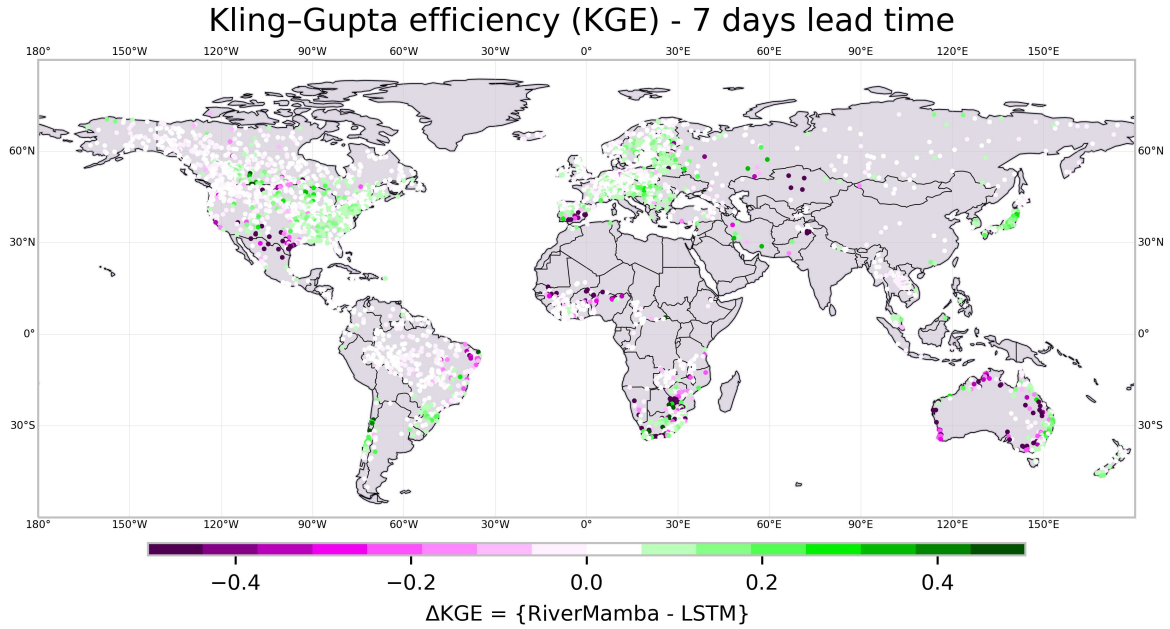


Figure 7.22: Comparison of KGE between RiverMamba and LSTM (Google) on GloFAS reanalysis (test set 2021-2024 temporally out-of-sample). RiverMamba improves over LSTM in 73% of the stations (P=3364).

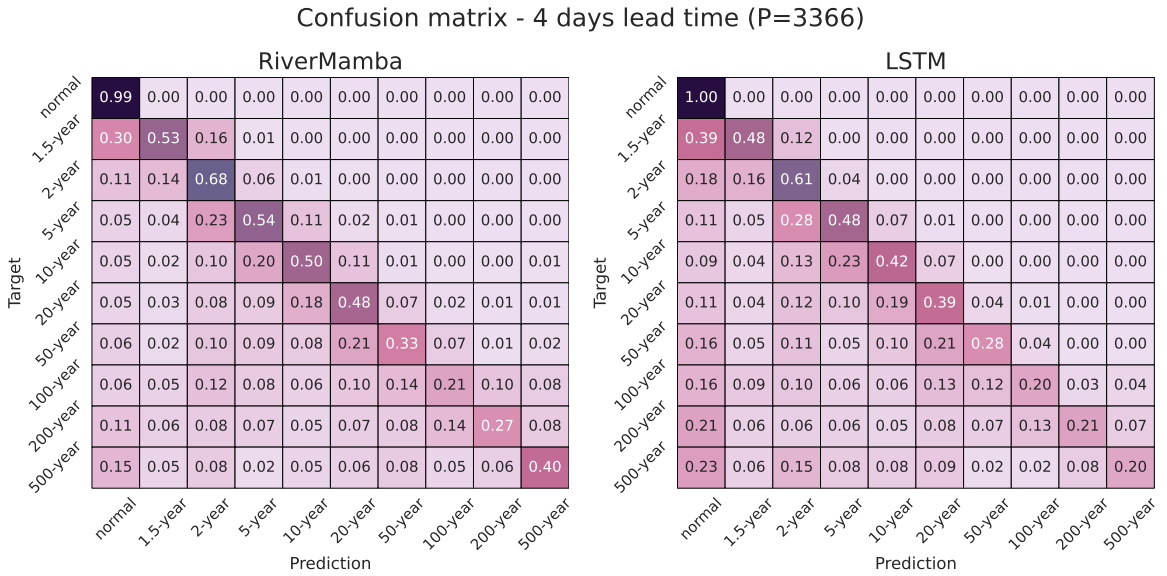


Figure 7.23: Comparison to LSTM on GloFAS reanalysis (test set 2021-2024 temporally out-of-sample).

7.3.3 Experiments on Gauged GRDC Observational River Discharge

Table 7.4 reports the performance on GRDC river discharge observations at gauged stations, which also includes the physics-based GloFAS reforecast model. As previously, Fig. 7.24 compares the forecast performance across multiple lead times and flood return periods. Compared to the results on GloFAS reanalysis (Table 7.3), all models show a noticeable drop in performance when evaluated on GRDC observations (Table 7.4). This decline likely stems from the fact that GloFAS simulates primarily naturalized discharge, with simplified representations of major reservoirs (*Harrigan et al.*, 2020; *Zajac et al.*, 2017), whereas GRDC reflects fully regulated flow, influenced by complex and unobserved human activities, such as dam operations and irrigation. This introduces biases that models cannot learn, especially in the absence of globally available data representing human water management, highlighting the challenge of predicting discharge under human-modified conditions.

The results show that traditional baselines such as Climatology and Persistence perform poorly. GloFAS performs much better than the baselines, but the R2 and KGE values are rather low due

	Validation (2019-2020)			Test (2021-2023)		
Model	R2 (↑)	KGE (↑)	F1 (↑)	R2 (↑)	KGE (↑)	F1 (↑)
Climatology	-0.0002	0.1342	–	-0.0013	0.0870	–
Persistence	0.1682	0.4569	0.1626	0.0660	0.3918	0.1462
GloFAS	0.3713	0.5412	0.2135	0.2892	0.4944	0.2044
LSTM	0.5437±0.0025	0.6572±0.0010	0.1724±0.0017	0.4615±0.0039	0.6141±0.0018	0.1475±0.0014
RiverMamba	**0.5943**±0.0016	**0.7015**±0.0007	**0.2577**±0.0046	**0.5057**±0.0028	**0.6612**±0.0010	**0.2427**±0.0111

Table 7.4: Results on GRDC gauged stations. (±) denotes the standard deviation for 3 runs. F1-score (↑) is averaged for all lead times and 1.5-20 year return periods.

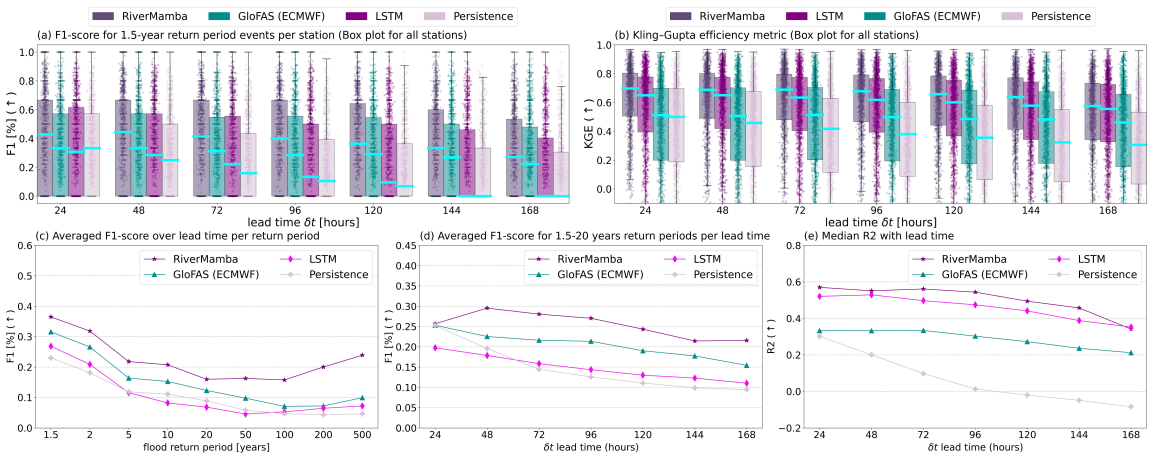


Figure 7.24: Results on gauged GRDC observations across lead times and flood return periods.

to the mentioned differences of physics-based models and observations. RiverMamba consistently outperforms the other methods for all metrics. Notably, RiverMamba shows less degradation in F1-score with increasing lead time, highlighting its strength in medium-range flood forecasting.

In Figs. 7.25-7.39, we plot additional results for the experiments on GRDC observational river discharge. In Figs. 7.25-7.28, we report the results for MAE, RMSE, R2, and R metrics with lead time. As can be seen RiverMamba outperforms all baselines on observations across these metrics. GloFAS which is a physics based model suffers on the absolute comparison to observational data like GRDC. This is because the output of GloFAS is simulated river discharge and the output is usually compared to observations in a relative manner, e.g., by comparing the occurrence of flood events rather than matching the absolute value of river discharge. However, RiverMamba can capture better the absolute value of observed river discharge along with the occurrence of flood events. Figs. 7.29 and 7.30 show the confusion matrix for RiverMamba, LSTM, and GloFAS. As can be seen, RiverMamba improves the true positive for different return periods and it reduces the false negatives compared to both LSTM and GloFAS. In Figs. 7.31-7.36, we report the results of F1-score, Precision, and Recall metrics for different return periods and lead times. Compared to the results on reanalysis data (Figs. 7.15-7.20), we can see that all baselines have relatively lower perform on observations. However, RiverMamba shows less degradation with lead time and higher return periods. Finally, in Figs. 7.37-7.39, we compare the results between RiverMamba, LSTM, and GloFAS for F1-score and KGE metrics spatially. The result shows that RiverMamba improves F1-score over LSTM in 42% of the stations and improves KGE over LSTM in 68% of the stations which is consistent with the result on reanalysis data (Figs. 7.21 and 7.22). The result also shows that RiverMamba improves F1-score over GloFAS in 38% of the stations and being better or equally better in 73% of the stations.

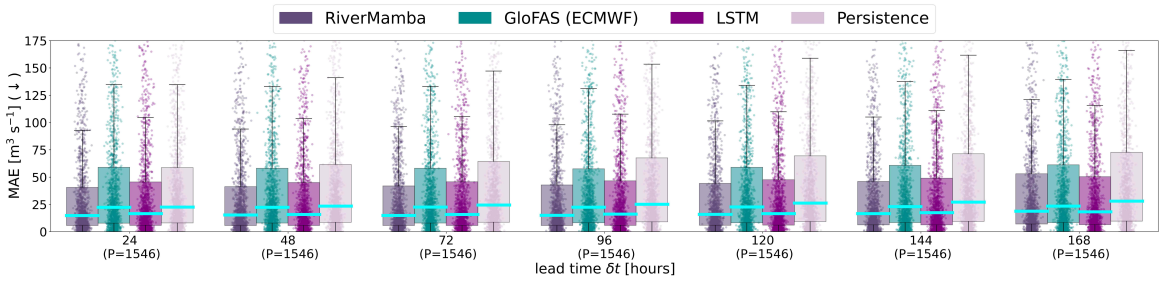


Figure 7.25: MAE of the river discharge forecasting with different lead time on GRDC observations (test set 2021-2023 temporally out-of-sample).

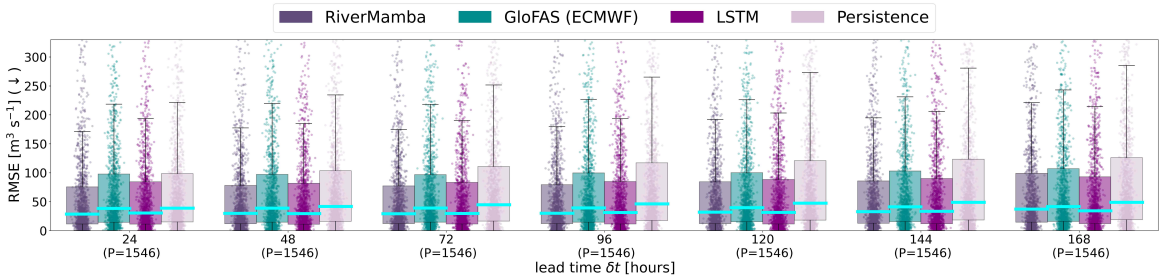


Figure 7.26: RMSE of the river discharge forecasting with different lead time on GRDC observations (test set 2021-2023 temporally out-of-sample).

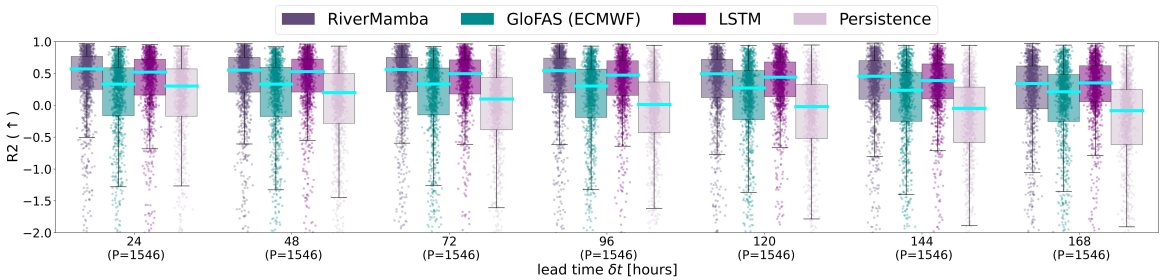


Figure 7.27: R2 (NSE) of the river discharge forecasting with different lead time on GRDC observations (test set 2021-2023 temporally out-of-sample).

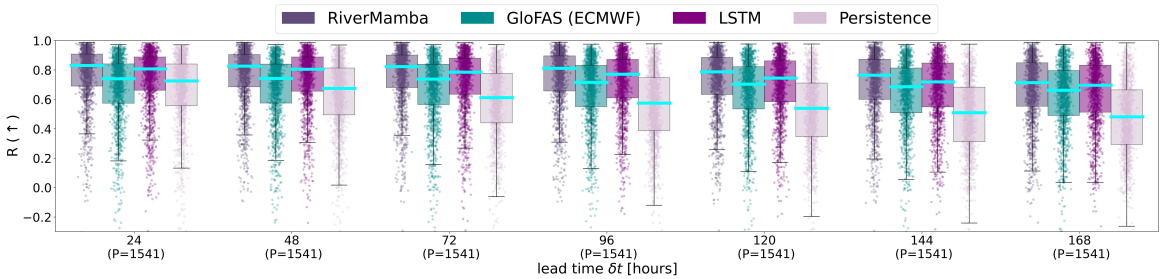


Figure 7.28: Pearson correlation (R) of the river discharge forecasting with different lead time on GRDC observations (test set 2021-2023 temporally out-of-sample).

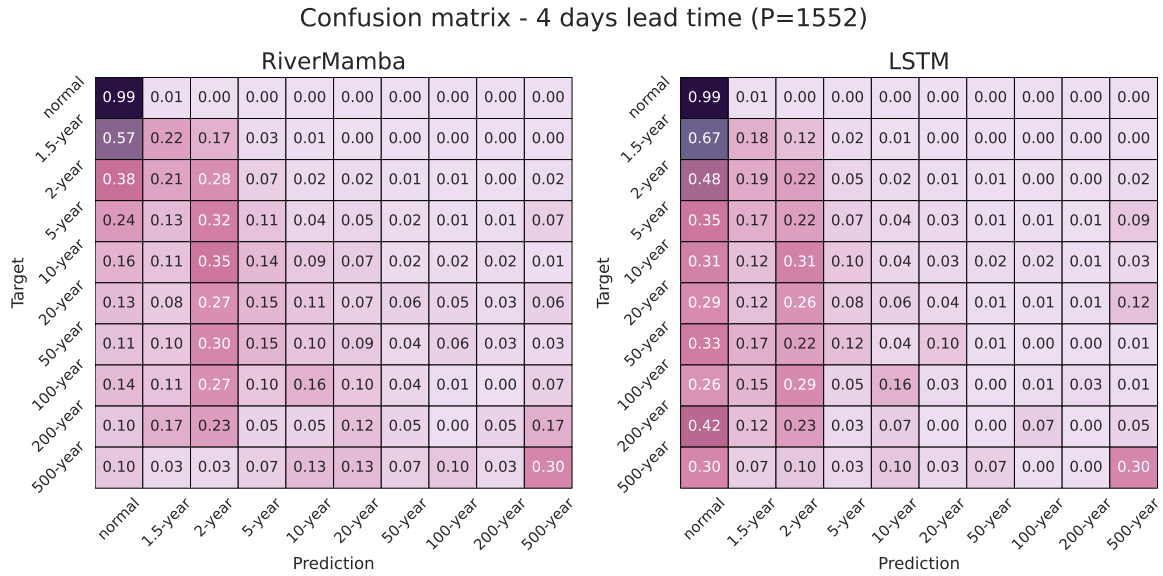


Figure 7.29: Comparison to LSTM on GRDC observations (test set 2021-2024 temporally out-of-sample).

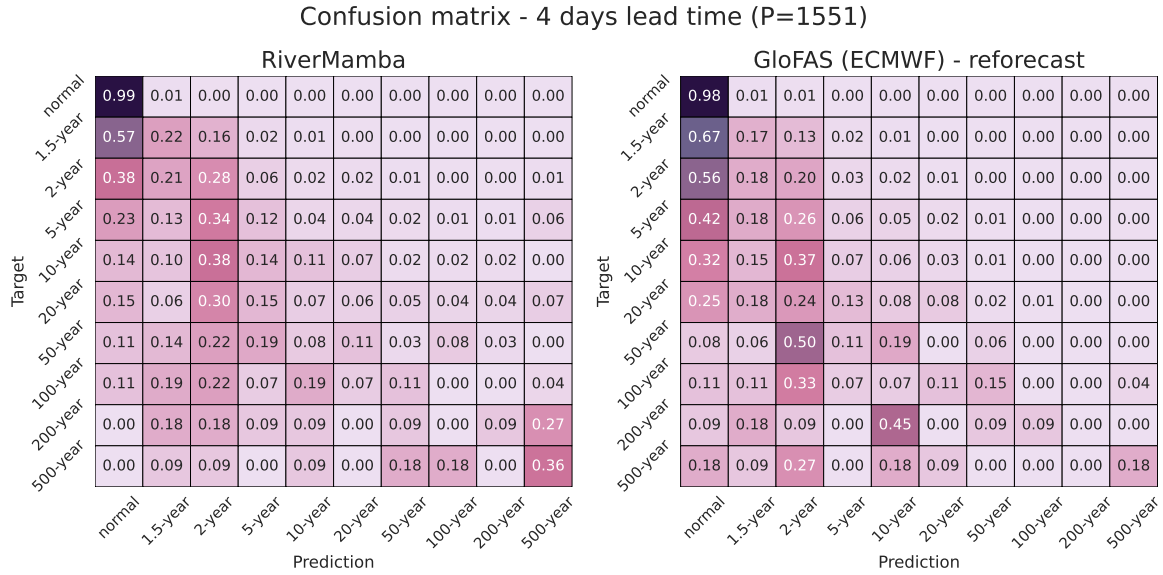


Figure 7.30: Comparison to GloFAS (ECMWF) - reforecast on GRDC observations (test set 2021-2023 temporally out-of-sample).

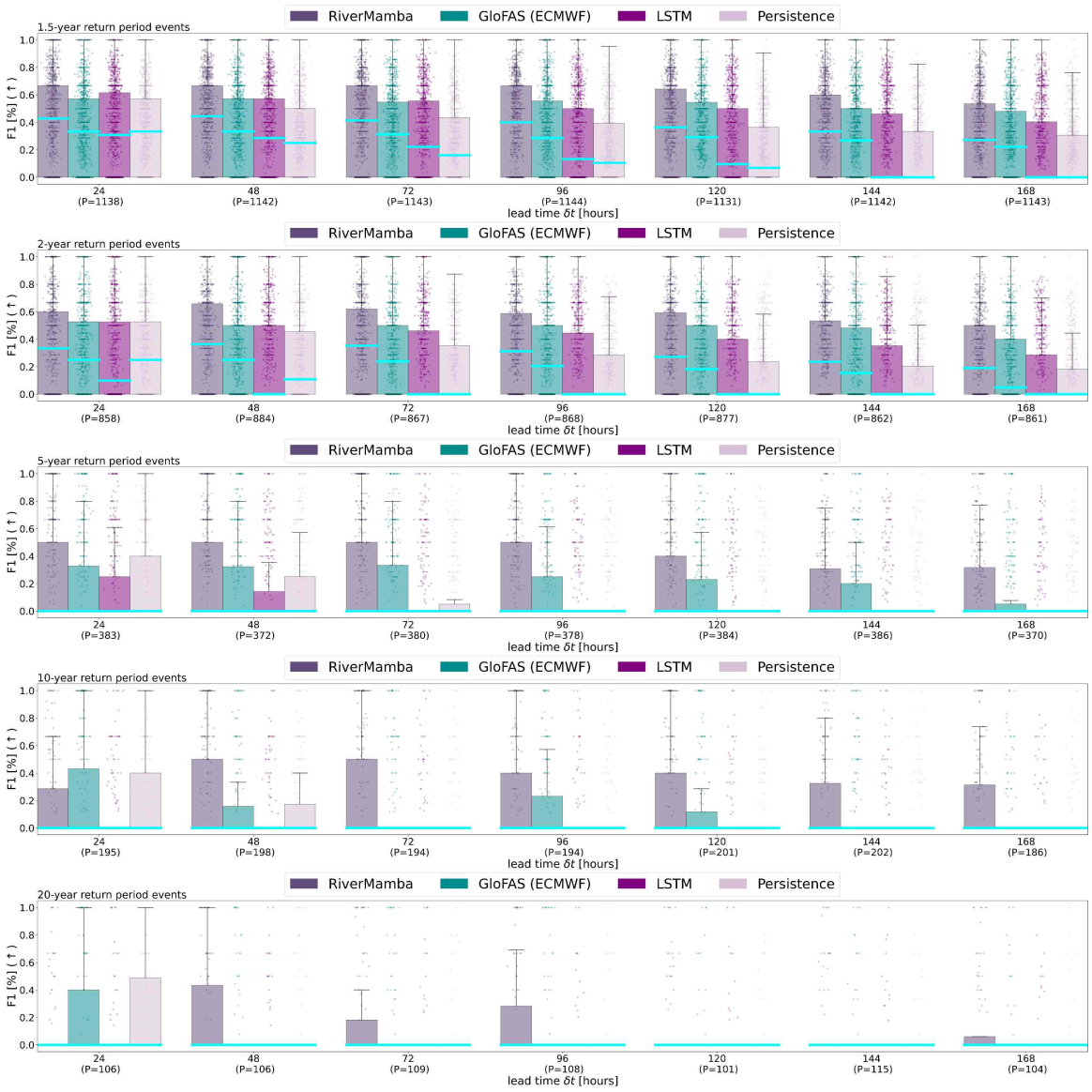


Figure 7.31: F1-score of flood forecasting for different return periods and lead time on GRDC observations (test set 2021-2023 temporally out-of-sample). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

Figure 7.32: Precision of flood forecasting for different return periods and lead time on GRDC observations (test set 2021-2023 temporally out-of-sample). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

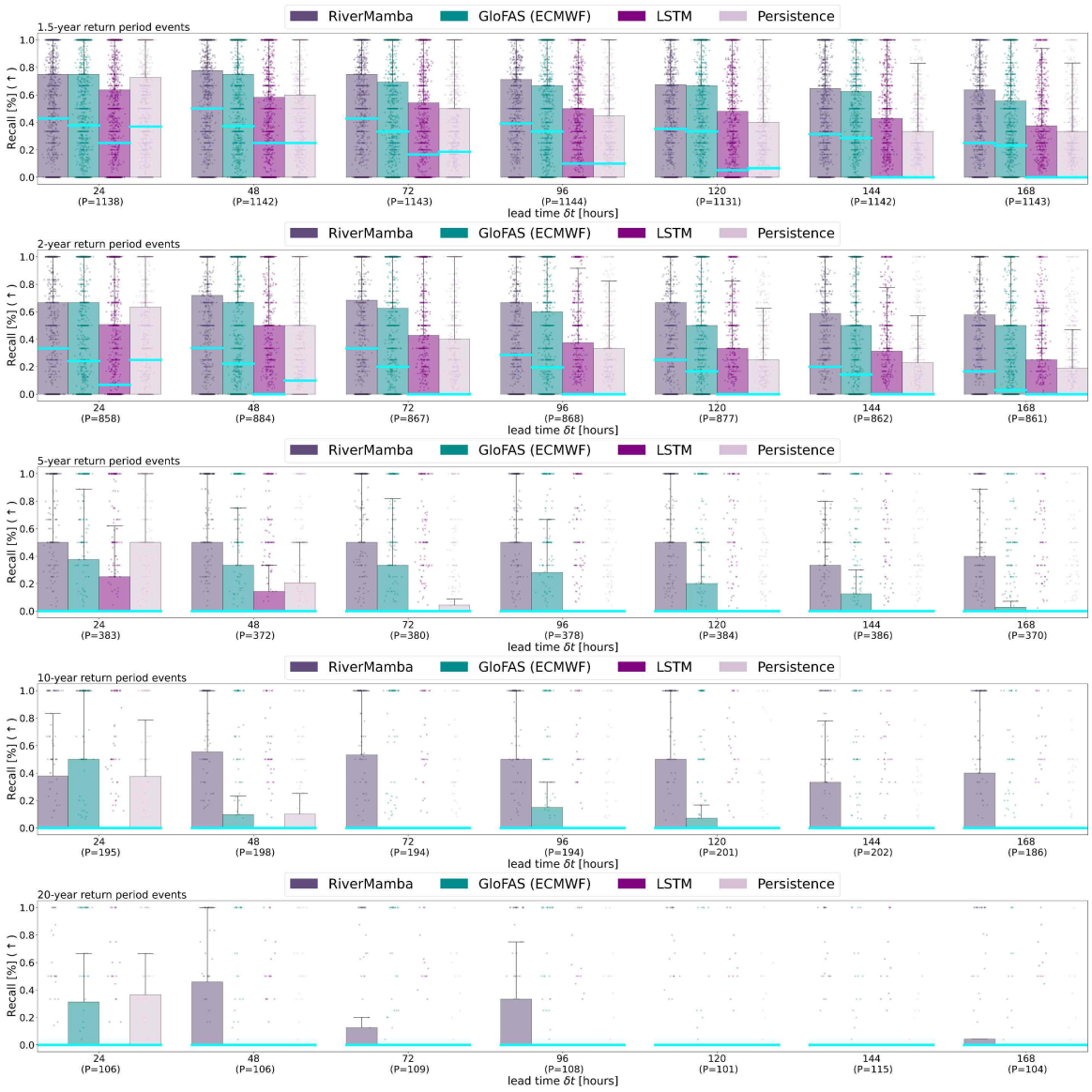


Figure 7.33: Recall of flood forecasting for different return periods and lead time on GRDC observations (test set 2021-2023 temporally out-of-sample). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

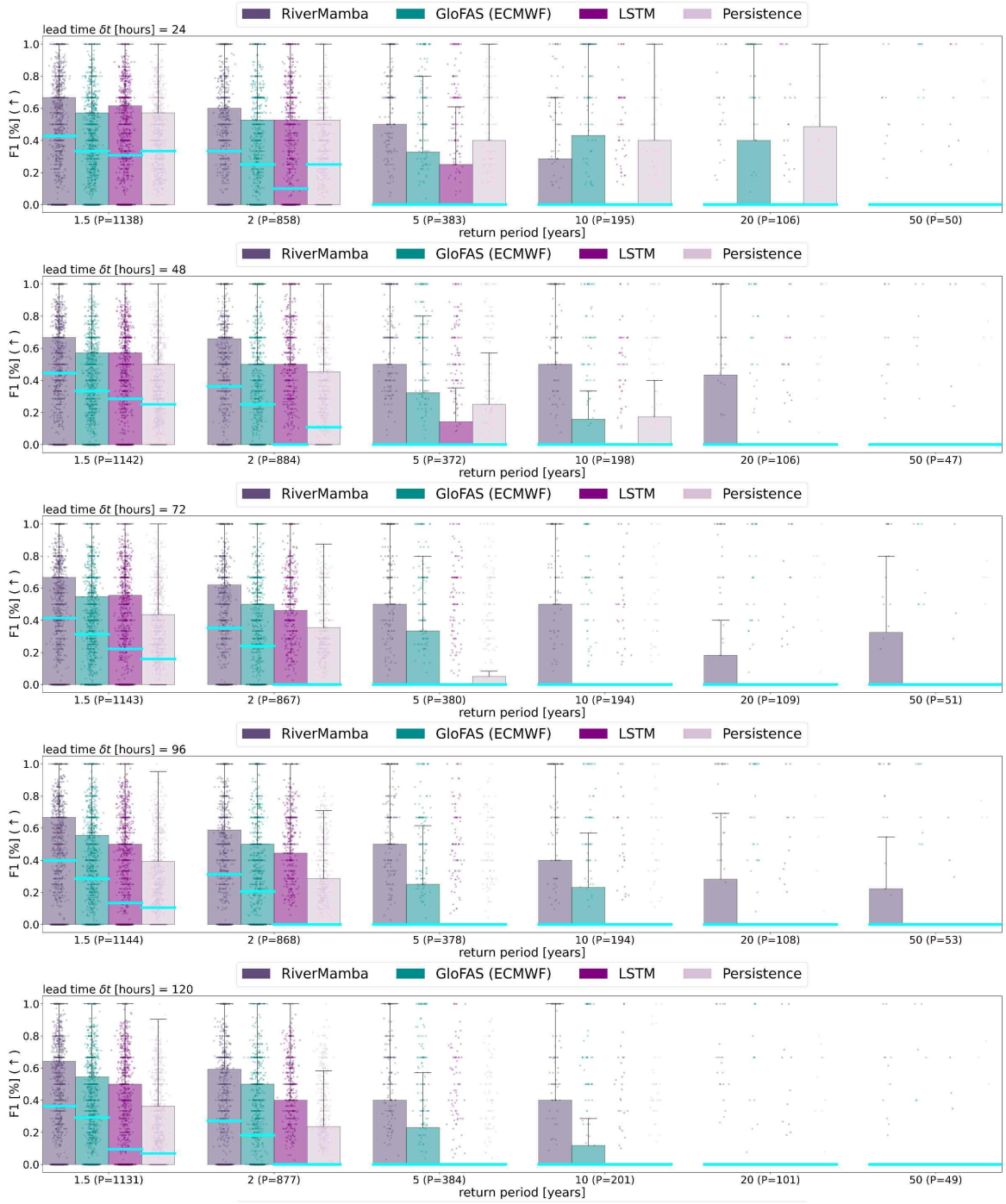


Figure 7.34: F1-score of flood forecasting for different lead time and return periods on GRDC observations (test set 2021-2023 temporally out-of-sample). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

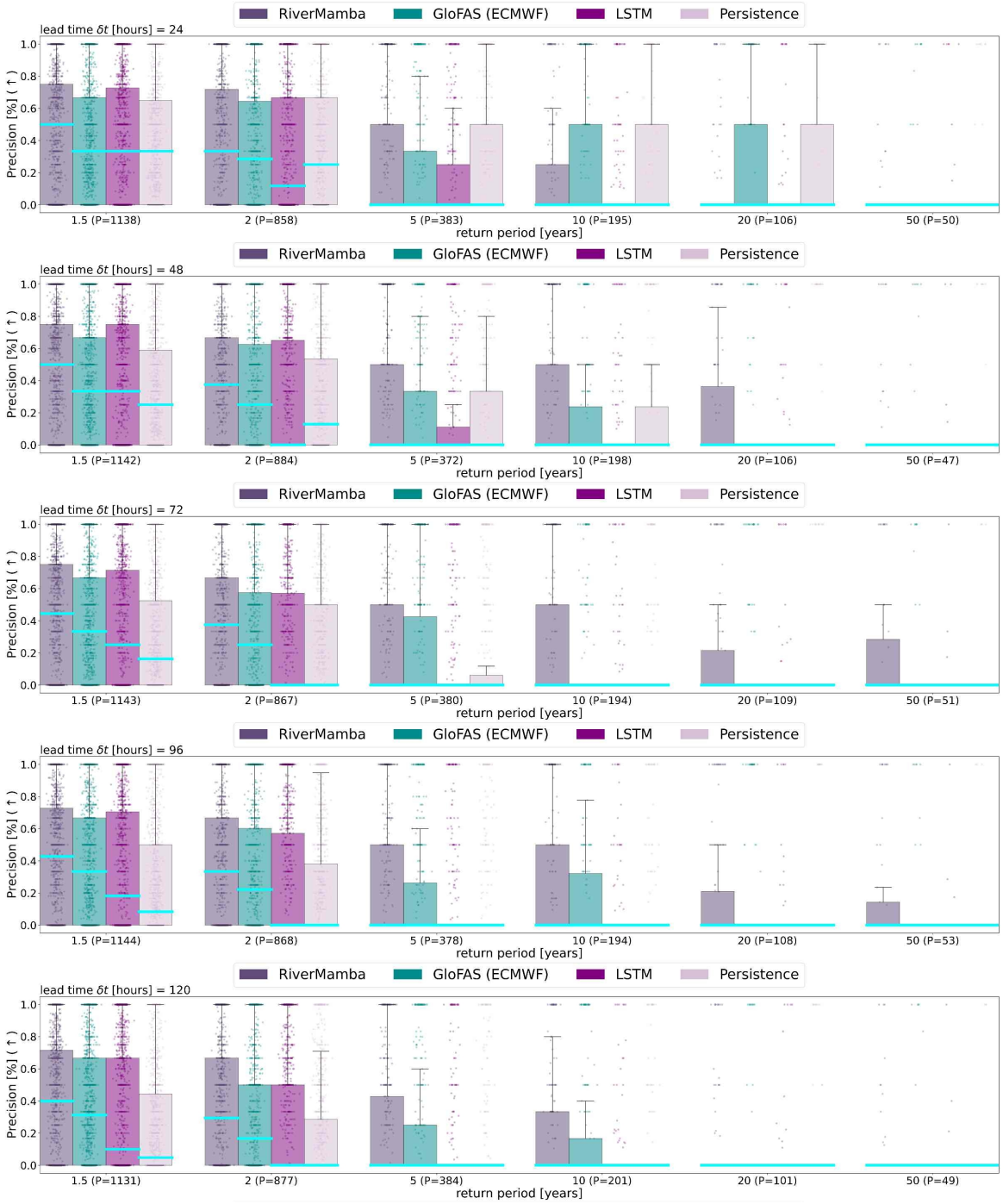


Figure 7.35: Precision of flood forecasting for different lead time and return periods on GRDC observations (test set 2021-2023 temporally out-of-sample). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

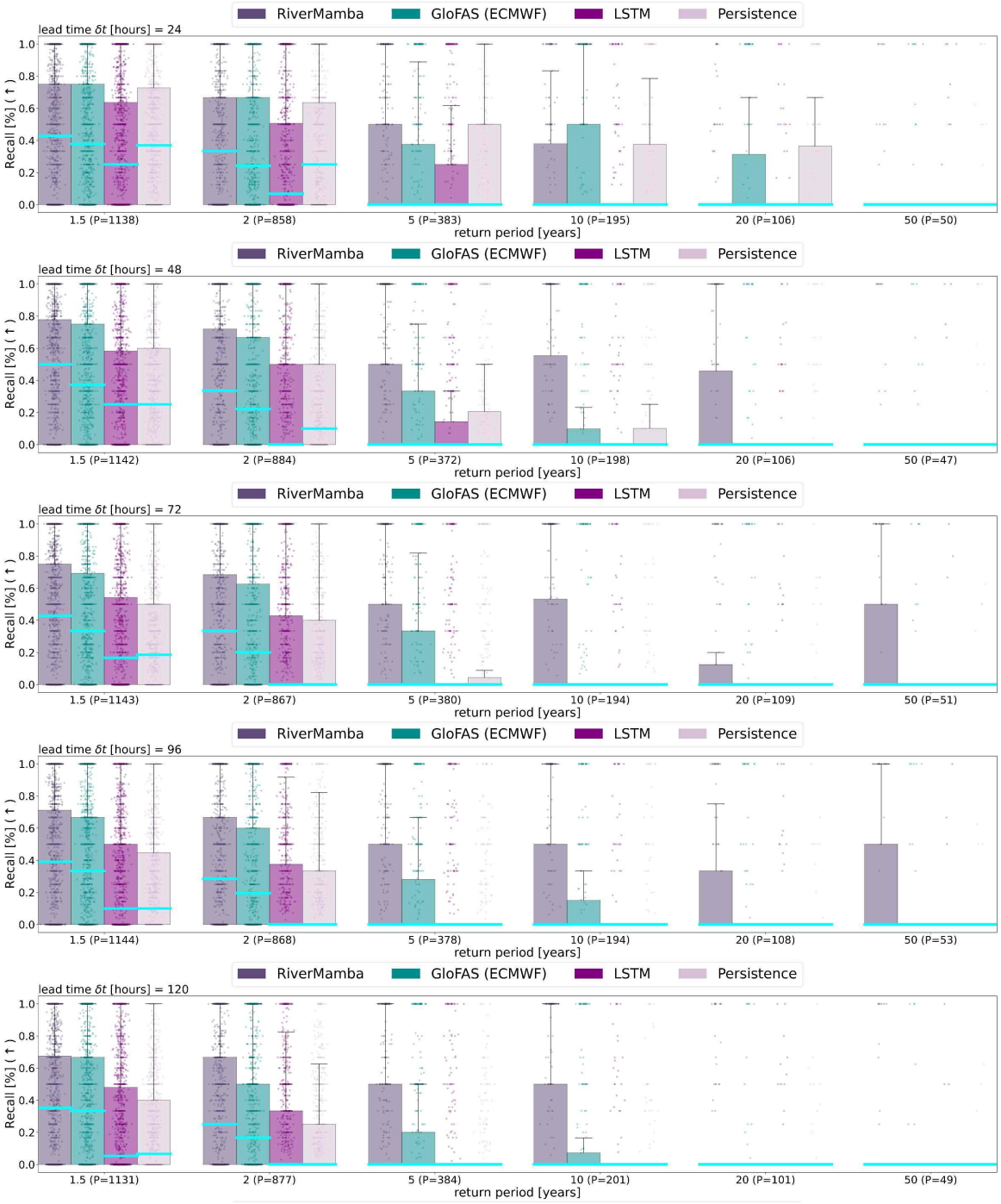


Figure 7.36: Recall of flood forecasting for different lead time and return periods on GRDC observations (test set 2021-2023 temporally out-of-sample). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

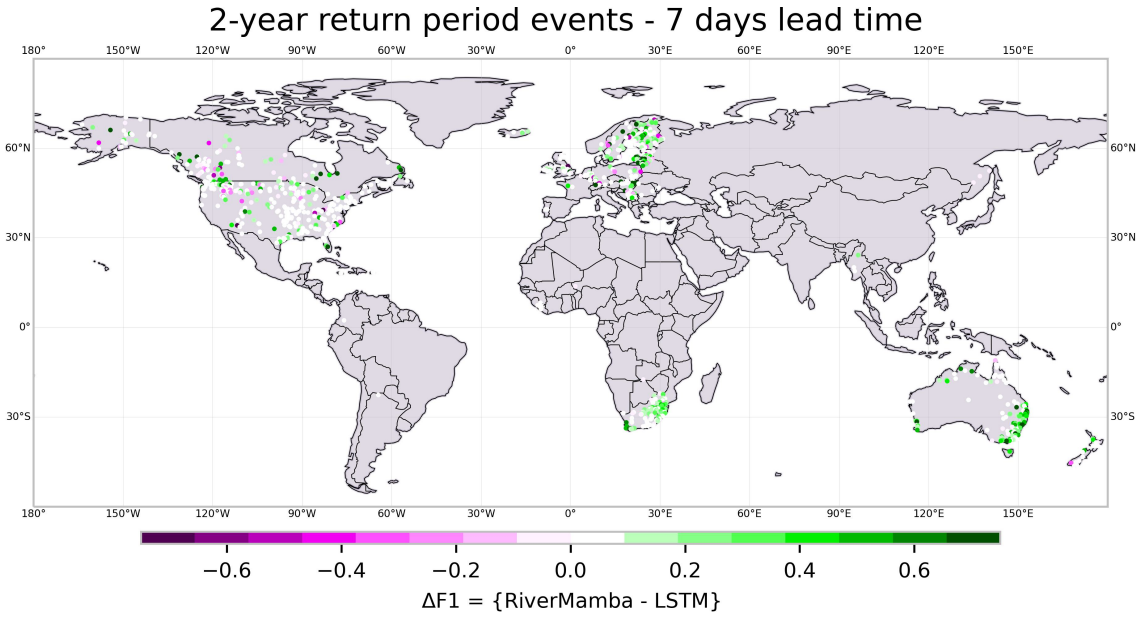


Figure 7.37: Comparison of F1-score between RiverMamba and LSTM (Google) on GRDC observations for the 2-year return period events (test set 2021-2023 temporally out-of-sample). RiverMamba improves over LSTM in 42% of the stations (P=861) and is better or equally better in 86% of the stations.

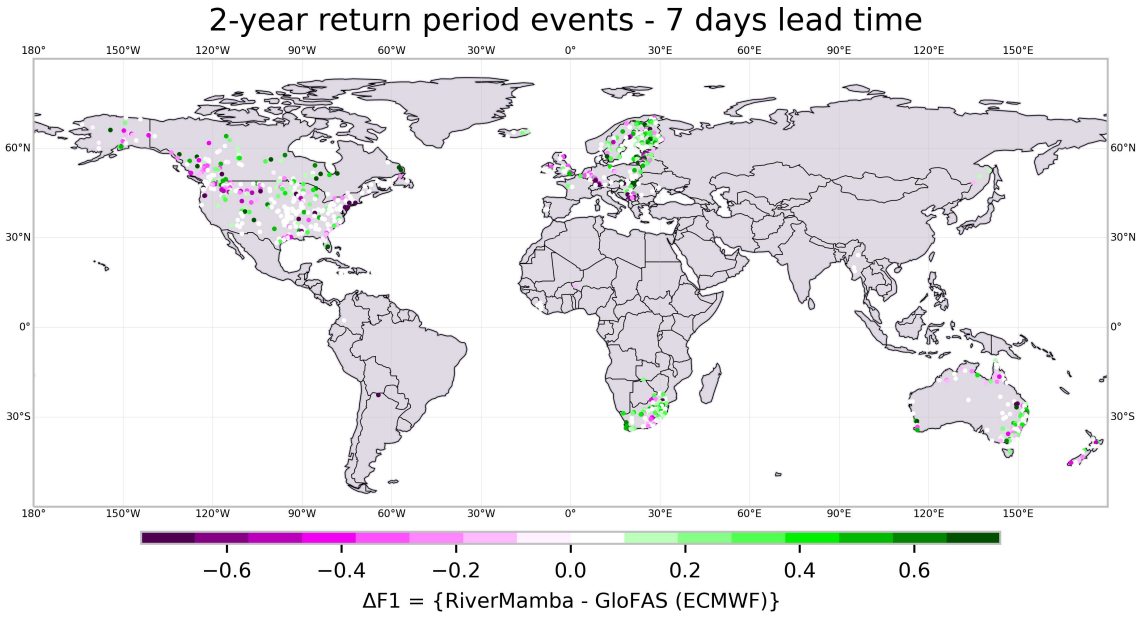


Figure 7.38: Comparison of F1-score between RiverMamba and GloFAS reforecast on GRDC observations for the 2-year return period events (test set 2021-2023 temporally out-of-sample). RiverMamba improves over GloFAS reforecast in 38% of the stations (P=861) and is better or equally better in 73% of the stations.

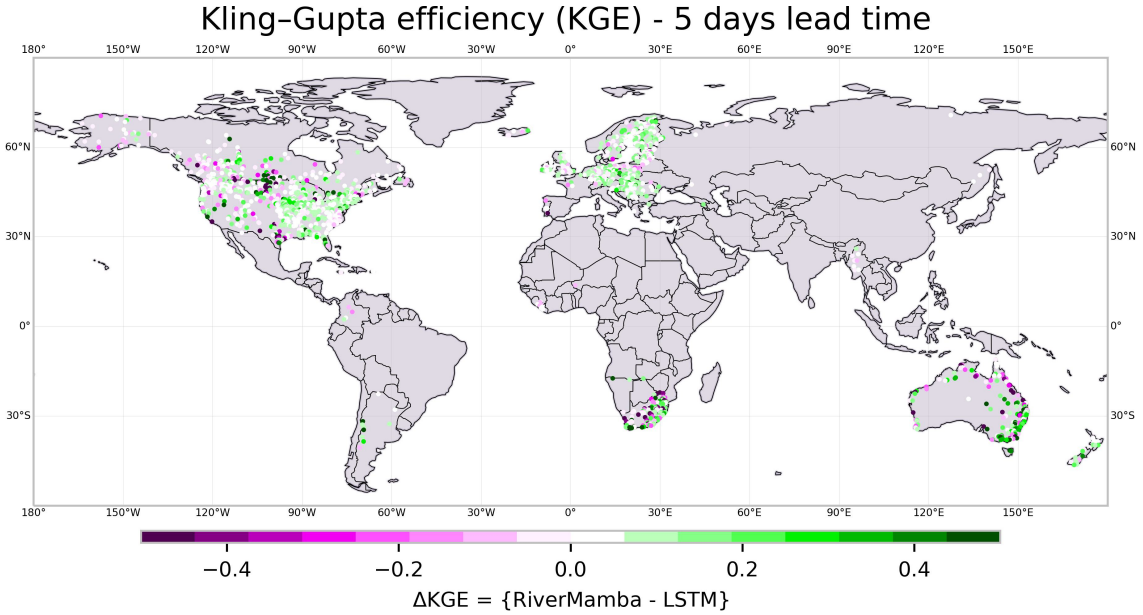


Figure 7.39: Comparison of KGE between RiverMamba and LSTM (Google) on GloFAS reanalysis (test set 2021-2023 temporally out-of-sample). RiverMamba improves over LSTM in 68% of the stations (P=1542).

7.3.4 Comparison with Google Reforecast on Ungauged GRDC

In this section, we evaluate RiverMamba against the published reforecast by *Nearing et al.* (2024) and available from *Nearing* (2023). For this, we split our data into 8 folds and evaluate on ungauged stations. All stations were predicted similar to *Nearing et al.* (2024) where each station was evaluated out-of-sample in both time and space. Note that the LSTM model for Google reforecast used more stations ($\sim$ 5680), while we used much less stations (3366). In addition, LSTM takes one year input as a hindcast, while RiverMamba takes only 4 days as input. Furthermore, RiverMamba does not use nowcasting data at time t and starts the input initial conditions at $t-1$ to mimic an operational forecast. There are also differences in the input initial conditions, e.g., LSTM uses precipitation estimates from the NASA Integrated Multi-satellite Retrievals for GPM (IMERG) early run as input. In addition, it uses HydroATLAS (*Linke et al.*, 2019) as geophysical and anthropogenic basin attributes. RiverMamba uses GloFAS reanalysis as an initial condition and LISFLOOD as static basin attributes.

Table 7.5 shows the overall performance for the years 2014-2021. The F1-score is averaged for all lead times and 1.5-20 year return periods. Fig. 7.41 (a) shows the averaged F1-score over lead time and Fig. 7.41 (b) shows the averaged F1-score over different return periods. As can be seen RiverMamba achieves good performance on ungauged setting. We expect that adding nowcasting (analysis data) and IMERG as input and an ensemble would improve the results of RiverMamba on ungauged basins further. The ungauged streamflow forecast becomes also better when the number of stations increases. More results are shown in Figs. 7.40 and 7.42. RiverMamba is more skillful than LSTM at longer lead time and higher return periods.

LSTM (Google reforecast from *Nearing et al.* (2024))	RiverMamba
0.2164	**0.2355**

Table 7.5: Comparison to Google reforecast on ungauged GRDC stations for the years 2014-2021. Shown is the averaged F1-score (↑) for all lead times and 1.5-20 year return periods.

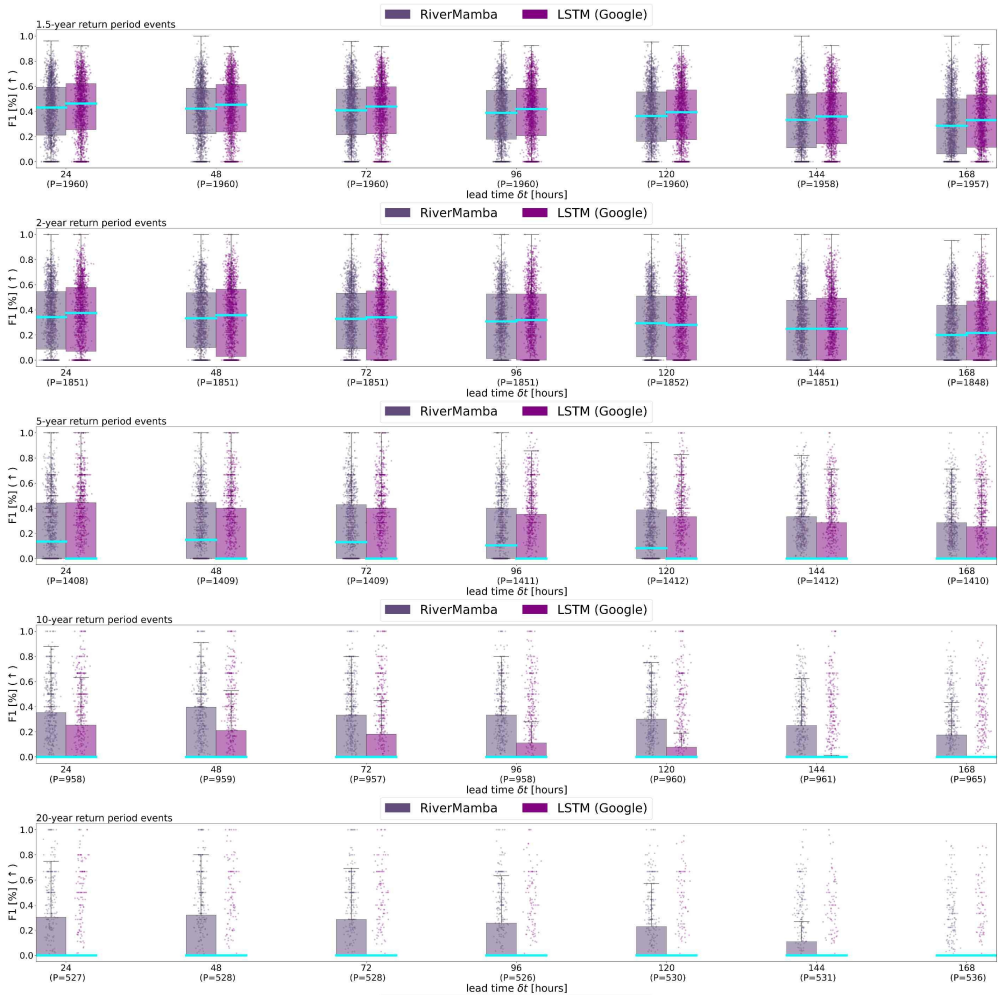


Figure 7.40: Comparison to Google reforecast. Shown is F1-score of flood forecasting for different return periods and lead time on ungauged GRDC stations (test set 2014-2021 out-of-sample in space and time). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

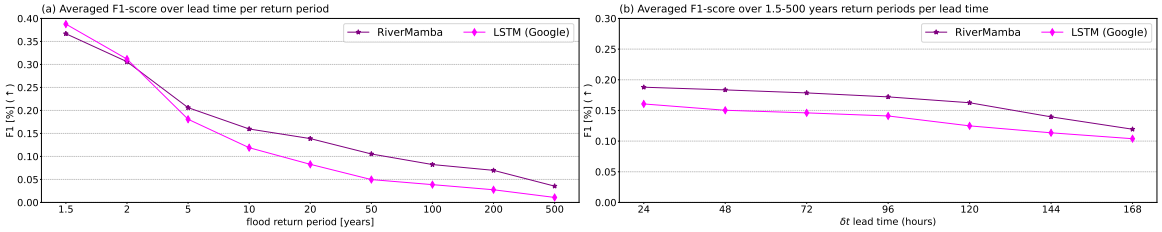


Figure 7.41: Comparison to Google reforecast on ungauged GRDC stations (test set 2014-2021 out-of-sample in space and time).

Figure 7.42: Comparison to Google reforecast. Shown is F1-score of flood forecasting for different lead time and return periods on ungauged GRDC stations (test set 2014-2021 out-of-sample in space and time). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

7.3.5 Comparison with Google Reforecast on Gauged GRDC

Similar to Section 7.3.4, here we compare to the published reforecast by *Nearing et al.* (2024) and available from *Nearing* (2023) but on gauged stations where all stations were evaluated out-of-sample in time for the years 2019-2021. For the differences between RiverMamba and the LSTM model by *Nearing et al.* (2024), see Section 7.3.4.

Table 7.6 shows the overall performance for the years 2019-2021. The F1-score is averaged for all lead times and 1.5-20 year return periods. More results are shown in Figs. 7.43-7.45. The results show that RiverMamba performs better than LSTM except for 1.5 return period events.

LSTM (Google reforecast from *Nearing et al.* (2024))	RiverMamba
0.2318	**0.2587**

Table 7.6: Comparison to Google reforecast on gauged GRDC stations for the years 2019-2021. Shown is the averaged F1-score (↑) for all lead times and 1.5-20 year return periods.

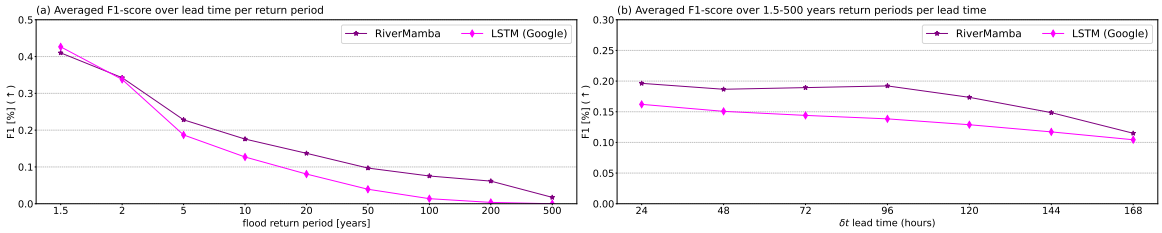


Figure 7.43: Comparison to Google reforecast on gauged GRDC stations (test set 2019-2021 out-of-sample in time).

7.3.6 Comparison to Operational GloFAS Forecsating on Gauged GRDC

Here, we compare to the archival operational forecast from GloFAS ECMWF (Section 7.3.1.3). This is different from the GloFAS reforecast and gives a more realistic assessment of the physics-based model. As can be seen, RiverMamba outperforms both the operational GloFAS and LSTM across all metrics.

Test (2023-2024)				
Model	MAE(↓)	R2(↑)	KGE(↑)	F1-score(↑)
GloFAS*	65.35	-0.0063	0.0439	0.1795
LSTM	53.07±0.39	-0.0005±0.0001	0.3147±0.0026	0.1120±0.0065
RiverMamba	**49.32**±0.18	**0.0001**±0.0001	**0.3821**±0.0076	**0.2358**±0.0211

*GloFAS operational forecast (*Harrigan et al.*, 2023).

Table 7.7: Results on GRDC gauged stations. (±) denotes the standard deviation for 3 runs. F1-score (↑) is averaged for all lead times and 1.5-20 year return periods.

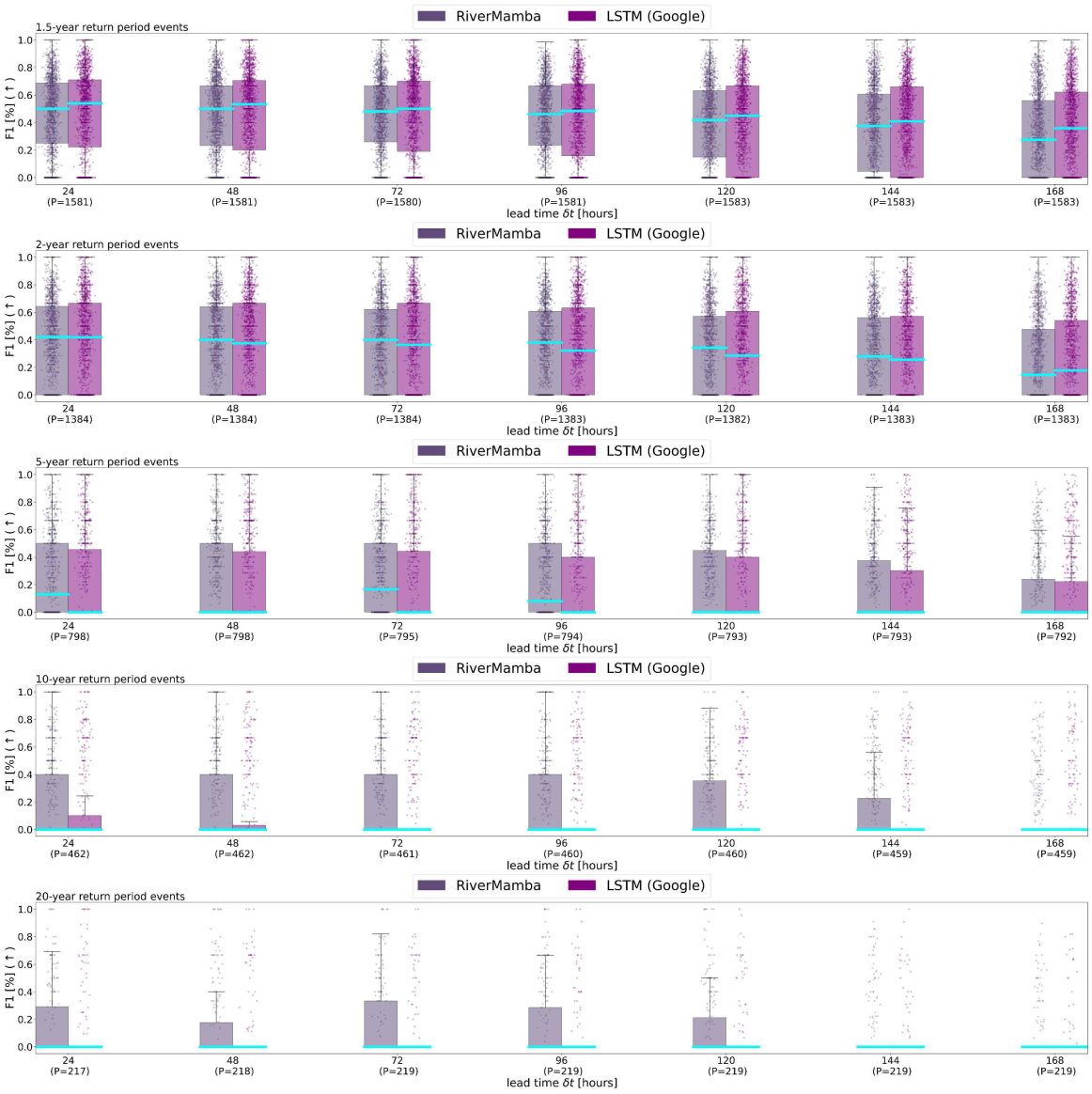


Figure 7.44: Comparison to Google reforecast. Shown is F1-score of flood forecasting for different return periods and lead time on gauged GRDC stations (test set 2019-2021 out-of-sample in time). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

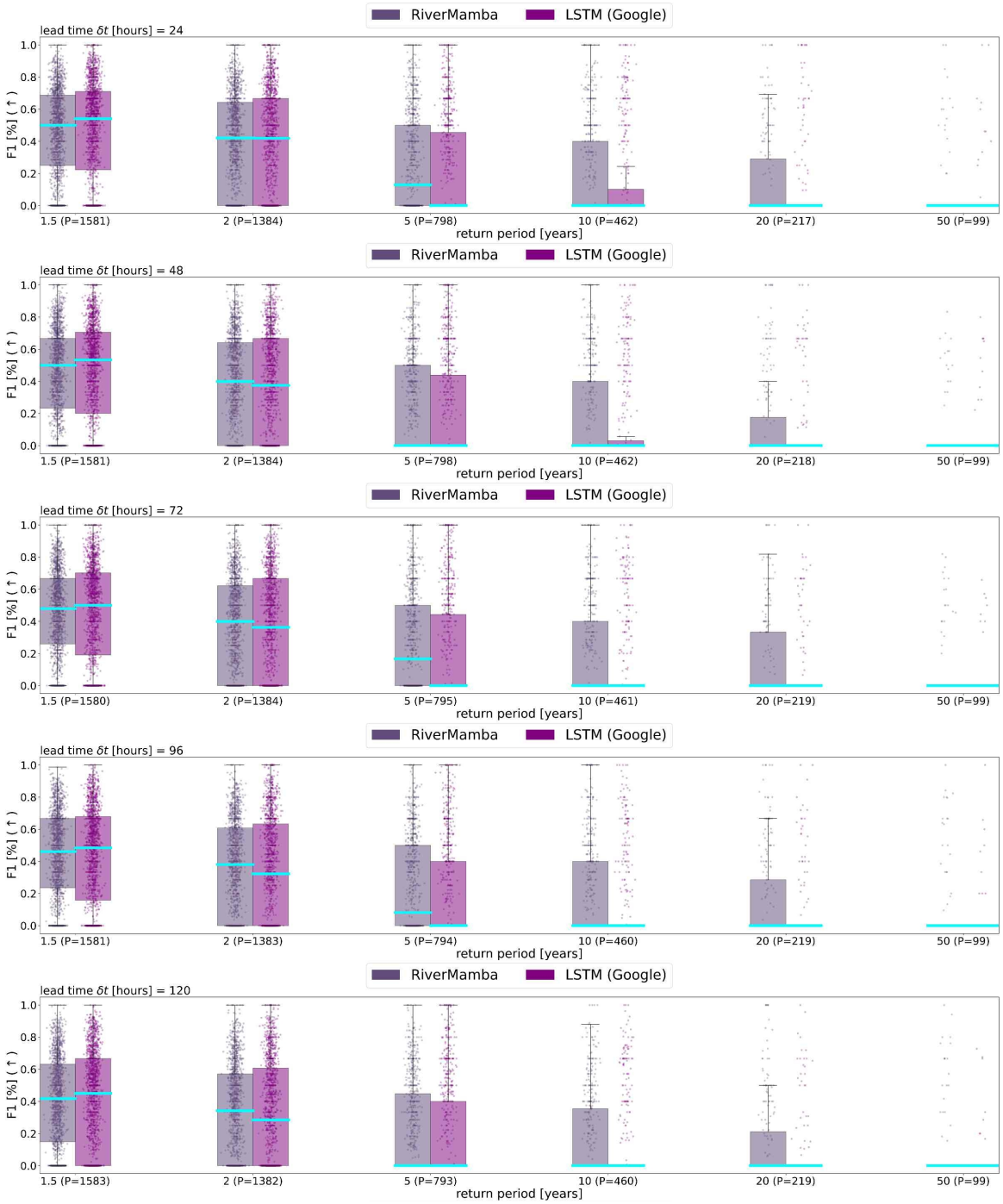


Figure 7.45: Comparison to Google reforecast. Shown is F1-score of flood forecasting for different lead time and return periods on gauged GRDC stations (test set 2019-2021 out-of-sample in time). Distribution quartiles are displayed in boxes, and the entire range excluding outliers is displayed in whiskers. The median score for the model is shown by the cyan line in the box.

7.3.7 Case Studies of Extreme Flood Events

7.3.7.1 2021 Western Europe flood

In this section, we present the daily river discharge at a gauged station on the Sauer River located at the border of Germany, France, and Luxembourg for the year 2021. Shown is the river discharge signal as predicted by RiverMamba, LSTM, and GloFAS reanalysis, and compared against GRDC observations. Particular attention is given to the extreme flood event in July 2021 (*Najafi et al.*, 2024; *Kracheletz et al.*, 2025), highlighted by the grey-shaded area in Fig. 7.46. To illustrate the meteorological drivers of this flood, we also show 7-day precipitation from ERA5 reanalysis and ECMWF HRES forecasts in Fig. 7.47.

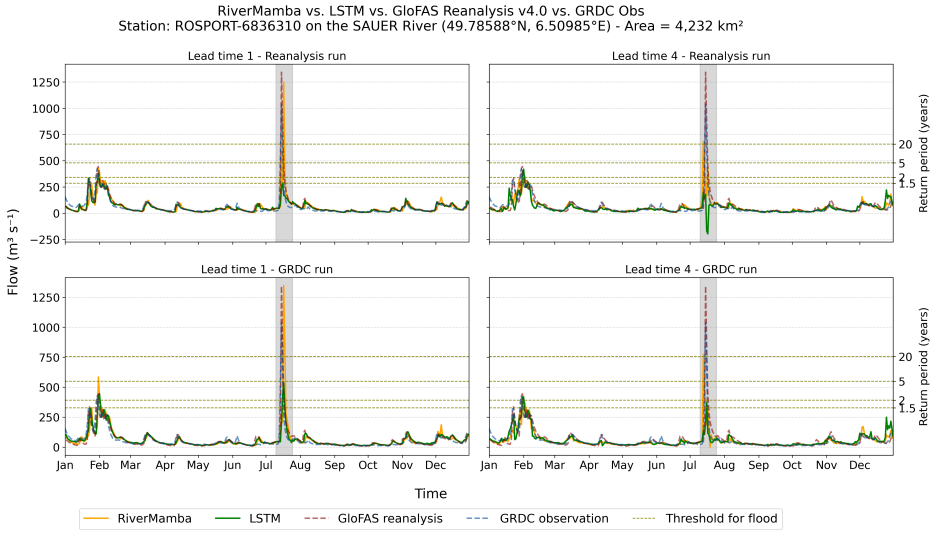


Figure 7.46: River discharge of the Sauer river in 2021 for RiverMamba (orange), LSTM (green), GloFAS reanalysis (dashed red), and GRDC observation (dashed green). The grey shaded area highlights the 2021 Germany flood between July 10 to July 20. The olive dashed lines represent the 1.5, 2, 5, and 20-year return periods calculated over reanalysis and GRDC observation data, respectively.

7.3.7.2 2024 Southeast Europe floods

In the following sections, we compare the daily flood severity map from GloFAS reanalysis as a ground truth and RiverMamba model at big flood events with different causes in 2024 and from different places around the Earth. These maps are usually used in the operational flood forecast service like GloFAS to provide a quick overview of the ongoing and upcoming flood events. As shown from these flood severity maps, RiverMamba can provide useful flood risk information at high spatial resolution to support decision-making. To the best of our knowledge, RiverMamba is the first AI model to demonstrate strong performance in predicting flood return periods globally under varying climate conditions at 5 km resolution across Europe, USA, Africa, and China. This further demonstrates its potential as a valuable component of operational flood early warning systems. It is important to note that the quality of the ECMWF HRES data driving the forecast is a key factor influencing RiverMamba performance.

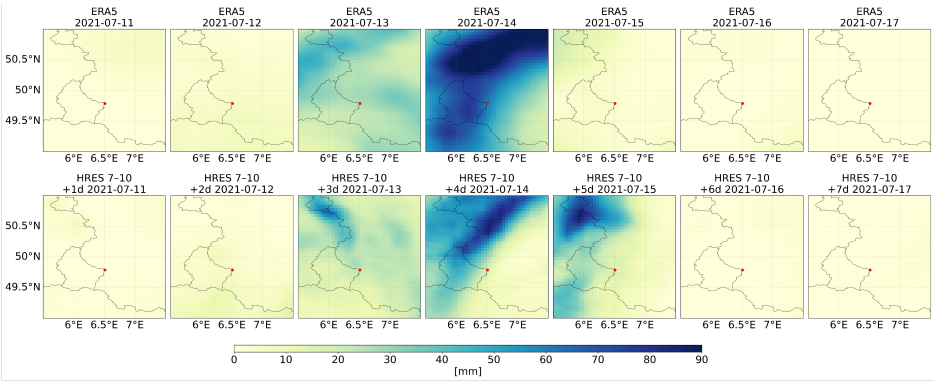


Figure 7.47: Daily total precipitation from 2021-07-11 to 2021-07-17 during the 2021 flood event in the target domain. First row shows precipitation as simulated by ERA5 Reanalysis and second row is the ECMWF HRES forecast issued at 2021-07-10 with 7-day lead time. The red dot is the location of the river discharge gauge station.

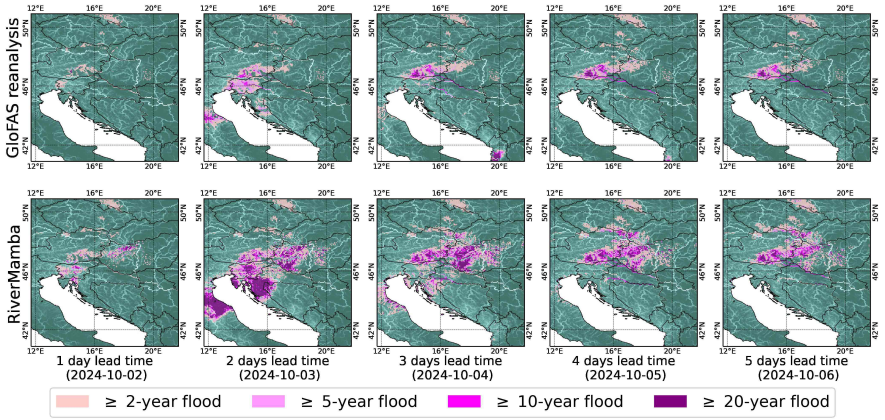


Figure 7.48: Comparison between GloFAS reanalysis (top, used as reference) and RiverMamba forecast (bottom) during the Southeast European flood in October 2024. Shown are flood severity maps at 5-km resolution where each panel shows flood extent at different lead times from 1 to 5 days.

7.3.7.3 2024 Central European floods

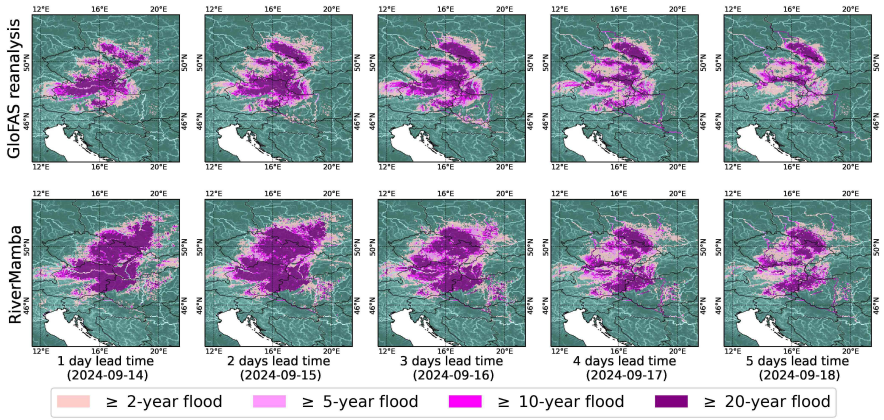


Figure 7.49: Comparison between GloFAS reanalysis (top, used as reference) and RiverMamba forecast (bottom) during the Central European floods event in 2024 *Hauer et al.* (2025). In September 2024, Storm Boris brought record-breaking rainfall to Central Europe, causing devastating floods across Austria, the Czech Republic, Poland, Romania, Slovakia, Germany, and Hungary. Studies indicate that climate change doubled the likelihood and increased the intensity of such extreme rainfall events, highlighting the growing impact of global warming on severe weather patterns. Shown are flood severity maps at 5-km resolution where each panel shows flood extent at different lead times from 1 to 5 days.

7.3.7.4 2024 Spanish floods

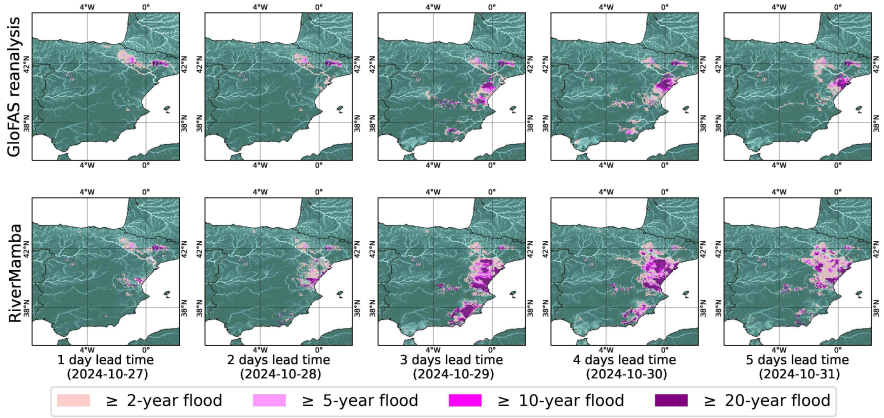


Figure 7.50: Comparison between GloFAS reanalysis (top, used as reference) and RiverMamba forecast (bottom) during the Spanish flood event in October 2024. The flood event primarily affecting the Valencia region, was caused by a cold drop (DANA) weather system intensified by climate change. Shown are flood severity maps at 5-km resolution where each panel shows flood extent at different lead times from 1 to 5 days.

7.3.7.5 2024 Saarland Germany flood

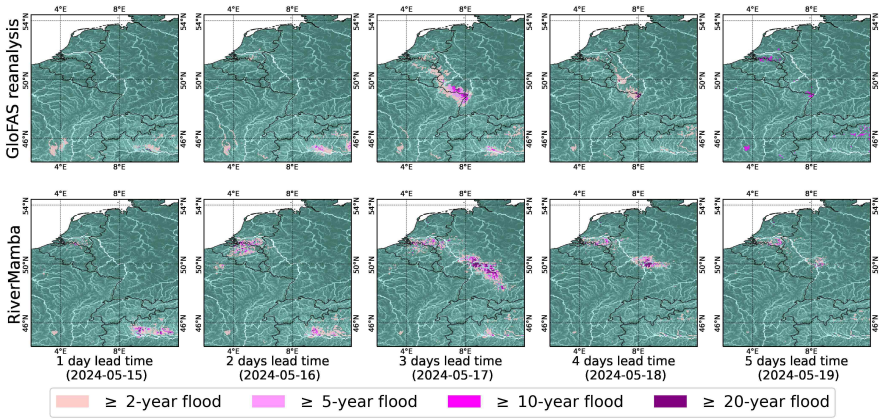


Figure 7.51: Comparison between GloFAS reanalysis (top, used as reference) and RiverMamba forecast (bottom) during the Saarland Germany flood event in May 2024. It was caused by thunderstorms and extreme rainfall, resulting in deadly floods and landslides across Saarland and Rheinland-Pfalz. However, RiverMamba predicted this flood event at a nearby location in Saarland, and the reason could be attributed to the inaccurate real-time ECMWF HRES forecast that drives the flood forecasting. Shown are flood severity maps at 5-km resolution where each panel shows flood extent at different lead times from 1 to 5 days.

7.3.7.6 2024 Kenya-Tanzania flood

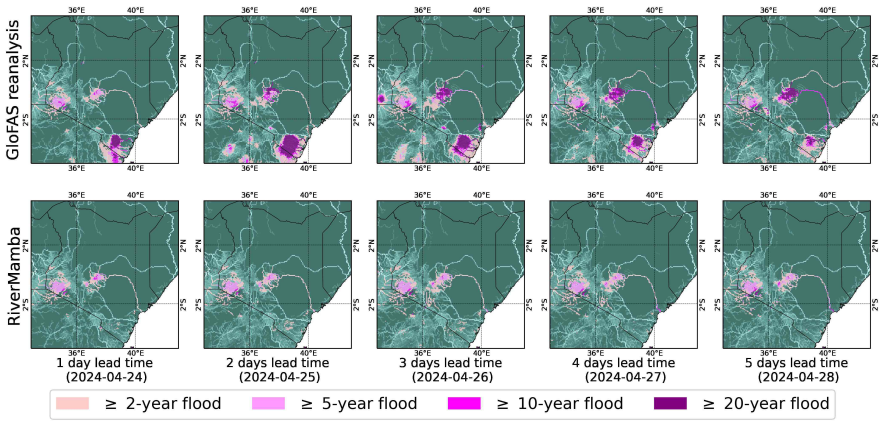


Figure 7.52: Comparison between GloFAS reanalysis (top, used as reference) and RiverMamba forecast (bottom) during the Kenya-Tanzania flood event in April 2024. This was the consequence of a combination of El Niño and a positive Indian Ocean Dipole, resulting in deaths, widespread displacement, and significant infrastructure damage. Shown are flood severity maps at 5-km resolution where each panel shows flood extent at different lead times from 1 to 5 days. Here points with less than 1 m^3/s discharge have been removed to erase artifacts on desert grids.

7.3.7.7 2024 California flood

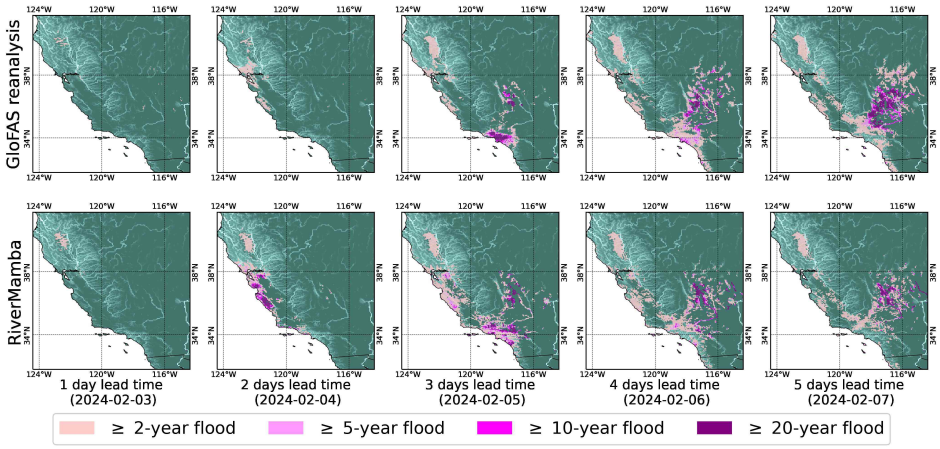


Figure 7.53: Comparison between GloFAS reanalysis (top, used as reference) and RiverMamba forecast (bottom) during the California, USA flood in February 2024. It was caused by two powerful atmospheric rivers. Shown are flood severity maps at 5-km resolution where each panel shows flood extent at different lead times from 1 to 5 days. Here points with less than 0.01 m^3/s discharge have been removed to erase artifacts on desert grids.

7.3.7.8 2024 Central-South China floods

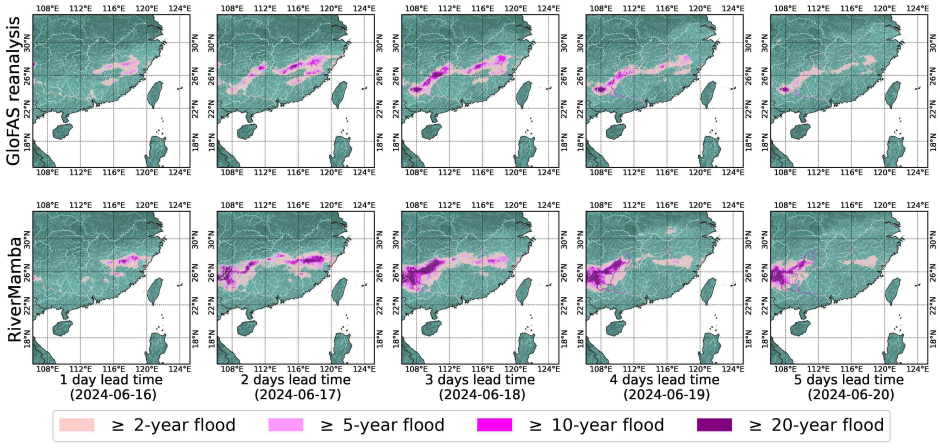


Figure 7.54: Comparison between GloFAS reanalysis (top, used as reference) and RiverMamba forecast (bottom) during the Central-South China flood event in June 2024 due to unprecedented rainfall. Shown are flood severity maps at 5-km resolution where each panel shows flood extent at different lead times from 1 to 5 days.

7.3.8 Ablation Studies

In the following, we discuss a set of ablation studies that are not performed globally but over Europe ($\phi = [30°\text{N} \rightarrow 60°\text{N}], \lambda = [10°\text{W} \rightarrow 40°\text{E}]$) which has $82,804$ points from the filtered diagnostic river points defined in Section 2.2.12 and includes 675 GRDC stations for evaluation.

7.3.8.1 Objective Function

$\hat{w}$	$\hat{u}$	KGE | F1 (↑)
✗	✗	0.9086 | 0.2236
✓	✗	0.9127 | 0.2859
✗	✓	0.9136 | 0.2593
✓	✓	**0.9205 | 0.2875**

Table 7.8: Ablation study on the loss function on the validation set over Europe.

In Table 7.8, we evaluate the impact of the weighting factor in the loss Eq. (7.9), which is based on $\hat{w}$ Eq. (7.10) and $\hat{u}$. The results show that both terms improve the results. $\hat{w}$ is important to focus on rare and more severe floods, increasing the F1 metric substantially (second row). $\hat{u}$ gives more weight to the forecast in the near future where $\mathbf{X}_{HRES}$ is more reliable, which is important due to the sequential structure of the forecast module. Using only $\hat{u}$ (third row) improves the results on both KGE and F1 metrics. Using both $\hat{w}$ and $\hat{u}$ (fourth row) gives the best results.

In Table 7.9, we study the effect of weighting floods not just by their return period but also by augmenting it with an additive flood offset of 1. To this end, we trained a model with weighted flood events by their return periods + flood offset of 1. The F1 score results are shown below in Table 7.9 for the reanalysis dataset and different return periods. Adding an offset of 1 does not improve the results.

Return period	1.5	2.0	5.0	10.0	20.0
Validation (2019-2020)					
W/ offset	0.4820	0.3760	0.2358	0.1790	0.1181
W/o offset	**0.4870**	**0.3767**	**0.2516**	**0.2015**	**0.1208**
Testing (2021-2024)					
W/ offset	0.6114	**0.5080**	**0.3125**	**0.2486**	0.1656
W/o offset	**0.6122**	0.5072	0.3031	0.2434	**0.1669**

Table 7.9: Ablation study on the validation set over Europe regarding the weighting in the objective function. Shown is F1-score for reanalysis data across different return periods.

7.3.8.2 Location Embedding

In Table 7.10, we show the benefit of using LOAN. In the first row, we duplicate the static features along the T dimension and concatenate them with the dynamic input. Using the LOAN layer in the

$\text{LOAN}_{(hindcast)}$	$\text{LOAN}_{(forecast)}$	KGE | F1 (↑)
✗	✗	0.9183 | 0.2790
✓	✗	0.9160 | 0.2827
✗	✓	0.9166 | 0.2931
✓	✓	**0.9205 | 0.2875**

Table 7.10: Ablation study on the location embedding on the validation set over Europe.

hindcast (second row) or forecast blocks (third row) increases the F1 score but decreases KGE. Using LOAN in both hindcast and forecast blocks balances the metrics (fourth row).

7.3.8.3 Forecasting Strategy

S-HRES	T-HRES	KGE | F1 (↑)
✗	✓	0.8862 | 0.2030
✓	✗	0.8869 | 0.2268
✓	✓	**0.9205 | 0.2875**

Table 7.11: Ablation study on the forecasting strategy in RiverMamba on the validation set over Europe.

Table 7.11 evaluates the impact of spatio-temporal modeling in the forecast module. In the first row, we remove the spatial relations in the forecast module by replacing the forecast blocks by point-wise MLPs. In this way, the data is processed after the last hindcast layer temporally but not spatially. This makes the model unaware of the spatial biases in the meteorological forcing $\mathbf{X}_{HRES}$. The second row denotes a setup where the forecast blocks do not get the features from the previous forecast block (Fig. 7.2) but directly from the last hindcast layer. In this case, we forecast river discharge for each lead time independently. The results show that in both cases the performance drops compared to our approach (third row), demonstrating the importance of spatio-temporal modeling.

7.3.8.4 Mamba vs. Transformer

In Fig. 7.55, we compare the model using Mamba blocks to a variant using Transformer blocks with Flash-Attention (*Dao et al.*, 2022; *Dao*, 2024). Both Mamba and Flash-Attention are efficient compared to a typical self-attention. However, Mamba scales better with the number of input tokens (Fig. 7.55 (left)), important for global modeling. The Transformer-based approach becomes computationally infeasible regarding the runtime for a larger number of input points. Both approaches have similar memory consumption which scales linearly with the sequence length. Using the bidirectional Mamba block increases the memory consumption slightly (Fig. 7.55 (right)).

Table 7.12 shows a comparison between Mamba and Flash-Attention regarding training time with different configurations. Flash-Attention uses local modeling (Fig.7.4 (b)). This requires rearranging the curves at each block, which becomes the main bottleneck for Flash-Attention compared to Mamba.

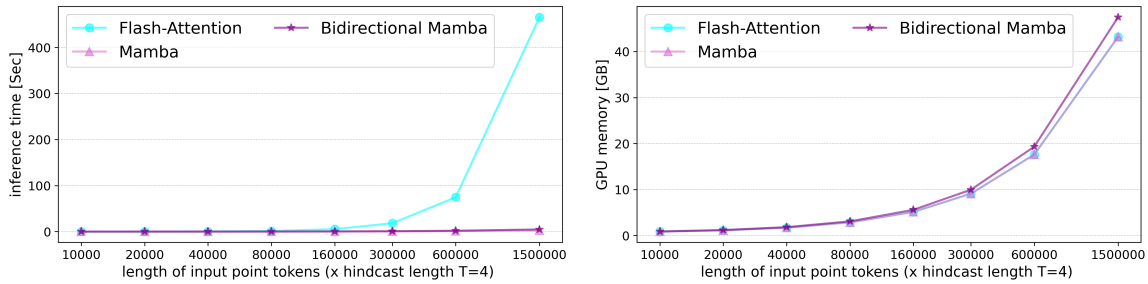


Figure 7.55: Comparison between Mamba and Transformer-based backbones.

Model	Time (min)	GPUs	CPUs
Flash-Attention	~ 145	$16 \times$ A100	64×4
Mamba	~ 82	$16 \times$ A100	64×4

Table 7.12: Training time on the reanalysis data for 1 epoch $(1,529,667$ points).

Table 7.13 shows the results for RiverMamba with different backbones. Flash-Attention and Mamba2 (*Dao and Gu*, 2024) achieve slightly lower performance compared to Mamba.

Backbone	# Params	KGE | F1 (↑)
Flash-Attention	5.03 M	0.9161 | 0.2804
Mamba	4.38 M	**0.9205** | **0.2875**
Mamba2	4.02 M	0.9169 | 0.2793

Table 7.13: Ablation studies for the RiverMamba backbone on the validation set over Europe.

7.3.8.5 Variable Importance

In Table 7.14, we study the performance of RiverMamba with different input features/variables. All input features have an impact on the performance. Removing the observational CPC data only slightly reduces the results for GloFAS reanalysis, but the decrease is larger for observational GRDC.

Removing GloFAS reanalysis and ERA5-Land initial conditions reduces the performance as well, but GloFAS reanalysis is more important. For GloFAS reanalysis (fourth row), we only drop GloFAS from the input and keep the data as a target to train the model. During inference, we still use the last step of GloFAS reanalysis and add it to $\Delta \mathbf{X}_{dis24}$ to generate the discharge. This ensures consistency within the table and isolates the impact of input GloFAS reanalysis on the model. RiverMamba still works without taking GloFAS as input. This highlights that the model is more than a post-processor of discharge data and can in fact be used as a backbone for hydrological modeling. Note that if we want to drop GloFAS completely, we need to change the objective function, e.g., by predicting the absolute value of river discharge or the change of discharge with respect to climatology. Finally, removing the meteorological forcing forecasts HRES has the biggest impact since weather forecasting is an important source of information. The best performance is achieved when we use all input variables.

Static (LISFLOOD)	CPC	ERA5	GloFAS-reanalysis	HRES	KGE | F1 (↑) (Reanalysis)	KGE | F1 (↑) (GRDC Obs)
✗	✓	✓	✓	✓	0.9091 | 0.2505	0.7633 | 0.1118
✓	✗	✓	✓	✓	0.9151 | 0.2842	0.7681 | 0.1102
✓	✓	✗	✓	✓	0.9077 | 0.2521	0.7731 | 0.1110
✓	✓	✓	✗	✓	0.9060 | 0.2450	0.7521 | 0.1226
✓	✓	✓	✓	✗	0.7972 | 0.1276	0.6757 | 0.0640
✓	✓	✓	✓	✓	**0.9205 | 0.2875**	**0.7838 | 0.1335**

Table 7.14: Ablation studies for variable importance on the validation set over Europe.

7.3.8.6 Pretraining on Reanalysis

In Table 7.15, we show the value of pretraining on GloFAS reanalysis data for the GRDC prediction. From our experiments, we can see a clear benefit of training on river discharge reanalysis before training the model on GRDC observations.

Pretrained on GloFAS reanalysis	KGE | F1 (↑)
✗	0.7406 | 0.0882
✓	**0.7838 | 0.1335**

Table 7.15: Ablation studies for pretraining on GloFAS reanalysis on the validation set over Europe.

7.3.8.7 Space-filling curves

In Table 7.16, we investigate the impact of various serialization patterns for RiverMamba. Our experiments show that sweep curves perform better than the other curves. Iterating between sweep and Gilbert curves (fourth column) improves the F1-score further. Iterating over all curves improves F1-score, but decreases KGE. For simplicity, we use thus the combination of sweep and Gilbert curves for our experiments.

Curve type	G	Z	S	S + G	S + G + Z
KGE | F1 (↑)	0.9156 | 0.2733	0.9156 | 0.2719	**0.9205** | 0.2826	**0.9205** | 0.2875	0.9163 | **0.2962**

Table 7.16: Ablation studies for space-filling curves on the validation set over Europe. The columns indicate the serialization patterns: G for Gilbert, Z for Zigzag, S for Sweep, Shuffle represents shuffling the order inside the hindcast layers. S and Z curves use both direction H and V and G uses both regular and trans Gilbert versions.

In Table 7.17, we remove the spatial modeling in RiverMamba completely and do the scanning only along the temporal dimension (first row). The results show the importance of spatio-temporal modeling.

The design of the scanning also plays a role. From Table 7.18 (a), we found that sequential scanning along the spatial dimension (P) works better. This is represented as scanning from P to T

Temporal modeling	spatio-temporal modeling	KGE | F1 (↑)
✓	✗	0.8726 | 0.1952
✗	✓	**0.9205** | **0.2875**

Table 7.17: Ablation regarding the spatio-temporal modeling on the validation set over Europe.

(P→T). In other words, the points are connected over time by scanning at time step t and continuing the scan at $t + 1$. In the second case (T→P), each point will be scanned along the time dimension (T) and then connected to the next point along the spatial dimension P.

In Table 7.18 (b), we split the curve into local curves similar to PointTransformer (*Wu et al.*, 2024) (Fig. 7.4 (b)). Using a larger receptive field gives the model more capability to extract up- and downstream features and to model adjacent catchments. In addition, local modeling needs more computations along the network, e.g., sorting, resorting and padding. Finally, bidirectional Mamba (Table 7.18 (c)) collects information about the streamflow from both side of the curve thus covering the whole domain and achieving a better performance than unidirectional Mamba.

(a) Curve order

T→P	P→T	KGE | F1 (↑)
✓	✗	0.9153 | 0.2807
✗	✓	**0.9205** | **0.2875**

(b) Curve type

Local	Global	KGE | F1 (↑)
✓	✗	0.9164 | **0.2893**
✗	✓	**0.9205** | 0.2875

(c) Bidirectional Curve

SSM	Bi-SSM	KGE | F1 (↑)
✓	✗	0.9113 | 0.2482
✗	✓	**0.9205** | **0.2875**

Table 7.18: Ablation studies for scan patterns on the validation set over Europe.

7.3.8.8 Experiments on HydroRIVERS

In this section, we explore the performance of RiverMamba with HydroRIVERS static maps (Section 2.2.11).

In Table 7.19, we compare the LISFLOOD with the HydroRIVERS static maps for prediction on both GloFAS reanalysis and GRDC data. Using HydroRIVERS performs worse than using LISFLOOD static maps.

HydroRIVERS	LISFLOOD	KGE | F1 (↑) (Reanalysis)	KGE | F1 (↑) (GRDC Obs)
✗	✗	0.9091 | 0.2505	0.7633 | 0.1118
✓	✗	0.9174 | 0.2622	0.7406 | 0.1227
✗	✓	**0.9205** | **0.2875**	**0.7838** | 0.1335

Table 7.19: Ablation studies on the validation set over Europe.

7.3.8.9 Activation function in LOAN

We conducted an additional experiment where we replaced the activation function in the LOAN layer by ReLU activation. GELU (*Hendrycks and Gimpel*, 2016) avoids the dying ReLU problem

and improves optimization. As can be seen from Table 7.20, GELU performs slightly better.

Activation function	ReLU	GELU
KGE | F1 (↑)	0.9143 | 0.2833	**0.9205 | 0.2875**

Table 7.20: Ablation study on the validation set over Europe regarding the activation function in LOAN layers.

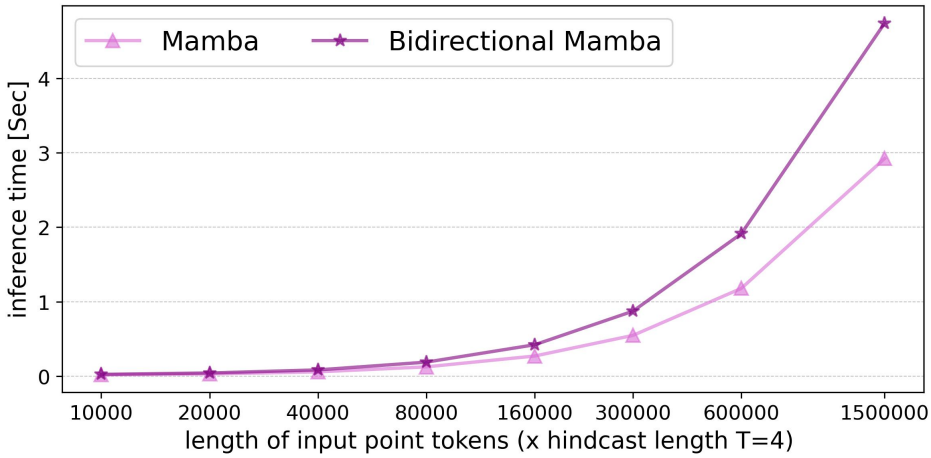


Figure 7.56: Inference time for RiverMamba with respect to the length of input tokens.

7.3.8.10 Computational Time

Neither Google (*Nearing et al.*, 2024) nor GloFAS (*Harrigan et al.*, 2023) provided the compute time for the operational forecast. The inference time for RiverMamba is reported in Fig. 7.56. In Table 7.21, we report the inference time (seconds) with respect to the number of input points for our model with 4 days as a hindcast (first row), a trained version of Google's LSTM with 4 days as a hindcast (second row), and a trained Google's LSTM version as in *Nearing et al.* (2024) with one year hincast (third row). We use one A100 GPU for all runs. All machine learning approaches are very fast. We expect that GloFAS is by several magnitudes slower, which is a practical advantage of machine learning approaches for this task.

Model	10K	20K	40K	80K	160K	300K	600K	1500K
RiverMamba (4 days hindcast)	0.026	0.044	0.086	0.190	0.423	0.874	1.914	4.739
LSTM (4 days hindcast)	0.005	0.009	0.015	0.027	0.053	0.098	-	-
LSTM (one year hindcast)	0.069	-	-	-	-	-	-	-

Table 7.21: Inference time in seconds.

7.4 Summary

In this chapter, we introduced RiverMamba, a novel deep learning approach for global, medium-range river discharge and flood forecasting. Due to its efficient structure and specialized scanning paths, RiverMamba maintains a very large receptive field, while scaling linearly with respect to the number of sampled points. As a result, RiverMamba is capable of forecasting high-resolution (0.05°) global river discharge maps. Further, the spatio-temporal modeling of the forecast blocks incorporates meteorological forcing and ensures a consistent forecast through space and time. Our analysis reveals that RiverMamba outperforms operational state-of-the-art deep learning and physics-based models on both reanalysis and observational data.

CHAPTER 8

Conclusion

In this chapter, we summarize the main contributions of this dissertation. Then, we highlight some limitations and challenges of our works in localizing and forecasting impacts of extreme weather and climate events. In the end of this chapter, we provide insights and promising future research directions.

Contents

8.1 Summary 211
8.2 Limitations 213
8.2.1 Wildfire Forecasting 213
8.2.2 Agricultural Drought Forecasting 214
8.2.3 Identifying Drivers of Extreme Events Impacts 214
8.2.4 Flood Forecasting 214
8.3 Outlook and Future Works 215
8.3.1 Ensemble Forecast 215
8.3.2 Forecasting Vegetation Conditions from Climate Scenarios 216
8.3.3 Agricultural Drought Forecasting 217
8.3.4 Identifying Drivers of Extreme Events Impacts 217
8.3.5 Flood Forecasting 218

8.1 Summary

In this dissertation, we presented unique approaches addressing the task of forecasting impacts of extreme weather and climate events. In addition, we introduced a novel deep learning method to identify drivers of extreme events impacts from reanalysis data.

In the context of wildfire forecasting, previous machine learning approaches duplicate static variables along the time dimension and give the static and dynamic variables simultaneously as an input to the model. In Chapter 4, we argued that this neglects the inherent differences between static and dynamic input variables. In contrast to previous lines of research, we have pursued the design of a deep learning model that takes only dynamic variables as the main input and adaptively normalizes the dynamic features inside the models based on their static features. To this end, we designed a module called Location-aware Adaptive Normalization (LOAN) that can be integrated into any existing neural network architecture with multi-modality (e.g., input from different sources). Inside the LOAN module, the dynamic features (e.g., temperature or pressure) are normalized or modulated based on the location attributes (e.g., elevation) in an end-to-end manner. We applied this idea for

wildfire forecasting where we developed a two-branch CNN that processes the spatial and spatio-temporal data in two separated 2D/3D CNN branches. By integrating the LOAN module into the developed model, we demonstrated an improvement compared to previous works and achieved state-of-the-art results on the FireCube dataset. We also applied the idea of adaptive normalization in Chapter 7 and showed the benefits of integrating LOAN for flood forecasting.

In Chapter 5, we addressed agricultural drought forecasting with deep learning. In this line of research, we investigated applying deep learning models to predict satellite-derived vegetation indices from a regional climate simulation. In other words, we used deep learning to predict how satellite images would look like according to climate scenarios. As deep learning models have difficulties in long-term forecasting of satellite images, we leveraged the climate simulation for long-term forecasting and deep learning to map the forecast variables into Normalized Difference Vegetation Index (NDVI) and Brightness Temperature (BT) images. These predicted images along with the climatology were then used to derive different vegetation and agricultural drought indices like Vegetation Health Index (VHI). For regional climate modeling, we used the Terrestrial Systems Modeling Platform (TSMP) simulation. We built our main deep learning model (Focal-TSMP) on focal modulation networks, extending their applications in climate science. The evaluation demonstrated that Focal-TSMP that was trained on satellite observations can predict NDVI and BT from a TSMP climate simulation with a sufficiently good alignment with satellite observations.

Afterwards, we addressed the task of identifying drivers of impacts extreme events. For instance, a major challenge in climate science is to understand the relations between impacts of extreme events (e.g., extreme agricultural droughts) and their drivers in climate data (e.g., anomalies in temperature and precipitation). What makes the task difficult is that the time delays between extreme impacts and their drivers vary seasonally and the drivers or anomalies can occur earlier in time and at a different location than the measured extreme event. In Chapter 6, we conducted experiments showing that, for this task, post-hoc explainable AI and typical anomaly detection algorithms like one-class, reconstruction-based, and multiple instance learning can result in spurious relations. Realizing this, we developed a novel deep learning model that is trained to predict spatio-temporal impacts of extremes. Instead of simply predicting the impacts of extremes, the model quantizes the spatio-temporal input multivariate variables into binary representation (driver or non-driver) and predicts the impacts of extremes only from the spatio-temporal binary or quantized maps. Like this, only the drivers in the input variables that are spatio-temporally correlated with the impacts of extreme are identified by the model. As an example of impact of extremes, we considered extreme agricultural droughts. We tested the model on two real-world reanalysis datasets and on diverse regions globally. Our proposed method showed a clear capability in identifying drivers from reanalysis data outperforming other typical methods. In addition to our experiments on real-world datasets, we proposed a novel framework for generating synthetic datasets. These synthetic datasets can serve as a benchmark for spatio-temporal anomaly detection tasks and for discoveries in climate research.

Our final contribution was related to river discharge and flood forecasting. Machine learning approaches in hydrology are largely steered toward local-scale applications. They are commonly based on catchment concept or lumped approaches, since distributed hydrological modeling with machine learning is still challenging. This is because the model has to consider the heterogeneous hydrological conditions and river routing inside the basins rather than treating the entire basin as a uniform lumped model. Realizing this, in Chapter 7, we framed the streamflow and flood forecasting as a vision task. We built a novel deep learning model on selective state space, an efficient alternative to

transformers, called RiverMamba, for global river discharge and flood forecasting. Notably, RiverMamba is the first deep learning approach that is capable of providing maps of global river discharge forecasting at 0.05°, and it introduces a novel methodology to hydrology. It is able to integrate sparse gauged observations, river attributes, high-resolution reanalysis data, and weather forecast. The efficient structure allows to model spatio-temporal relations at very large spatio-temporal context. We evaluate RiverMamba on both long-term real-world reanalysis and observational data. Throughout an extensive evaluation, we have showed that, for medium-range and global flood forecasting, RiverMamba outperforms state of the art models like Google's LSTM and the physics-based Global Flood Awareness System (GloFAS) operated by ECMWF.

Depending on the task and the scale of the dataset, different architectures can be recommended. For example, for short-term forecasting at a local scale with relatively small dataset (Chapter 4), a CNN is still a good choice compared to a vanilla transformer backbone. This is probably because vanilla transformers are prone to overfitting on such a dataset. For a prediction task from a climate simulation, and given the promising results in Chapter 5, a focal modulation backbone is encouraged over a vanilla transformer backbone for feature extraction. Focal modulation requires less computation and has more inductive bias than transformers. Moreover, Mamba can handle sparse input, in contrast to focal modulation. In addition, it can handle long sequences more efficiently than Flash-attention. When dealing with a potentially sparse map and for medium-range forecast at a large scale, such as river discharge forecasting (Chapter 7), using a model with linear computational complexity with respect to the input tokens like Mamba, is recommended. Finally, for the task of identifying anomalies and drivers from multivariate climate data (Chapter 6), it is important to design tailored methods for climate data and to check for spurious relations that can result from post-hoc explainable AI and typical anomaly detection algorithms.

8.2 Limitations

In this section, we discuss and highlight some limitations related to the works we have introduced in this dissertation. Such limitations should be taken into consideration, e.g., when applying the developed approaches and also when interpreting the results obtained by the developed models. As the research field of intersecting computer vision and climate science continues to evolve, we are confident that future efforts will overcome most of these limitations.

8.2.1 Wildfire Forecasting

In Chapter 4, we demonstrated that a two-branches CNN model with adaptive normalization and time encoding is a reasonable choice for wildfire forecasting. Although our approach demonstrated a substantial improvement compared to previous works for wildfire forecasting, it still has some limitations. Despite the fact that our framework includes domain knowledge through the normalization layer and absolute time encoding, it does not incorporate physical knowledge about the Earth system. Furthermore, the FireCube dataset covers only parts of Eastern Mediterranean and the years 2009-2021. There is a need for more standardized datasets for wildfire forecasting at a continental scale and longer time periods. Recently, benchmark datasets for wildfire risk forecasting have been released (*Kondylatos et al.*, 2023; *Gerard et al.*, 2023; *Karasante et al.*, 2025; *Xu et al.*, 2025). Our proposed approached can be tested on these datasets in the future. Moreover, the false positives are

also related to the resolution of the ERA5-Land data that have been used in the FireCube dataset. Higher resolution meteorological data can improve the performance. Finally, there may be hidden events that are correlated with climate variability and extreme weather conditions. It is an open question how these impact the forecast quality and if additional input variables will be needed to improve the forecast accuracy.

8.2.2 Agricultural Drought Forecasting

Our evaluation in Chapter 5 confirms that Focal-TSMP has the capacity to predict NDVI and BT from a TSMP climate simulation. However, there are still some limitations. First, there are uncertainties in the input simulation linked to potential errors in the driving forces and spin-up initialization. Second, vegetation products are derived from different sensors and their quality varies across regions. Third, predictions over regions with high latitude and with little seasonality, e.g., desert and high mountains have relatively higher errors.

Finally, there is a key limitation when applying Focal-TSMP for human intervention scenarios (e.g., land- and water use scenarios), which is the absence of observational ground truth data and benchmark to evaluate the model performance under such scenarios.

8.2.3 Identifying Drivers of Extreme Events Impacts

While we have shown that our approach in Chapter 6 for identifying drivers of extreme events impacts outperforms other approaches, the study has some limitations. First, evaluating ability to handle a very large number of climate variables in a unified model needs further examination. Similarly, performing an additional preprocessing of specific variables like accumulating precipitation over many weeks might also improve the results. Second, modeling the temporal relations is limited by the time window T. Moreover, teleconnections of climatic anomalies can occur in distant regions on Earth, e.g., affects of El Niño and La Niña variability on drought and flood (*Das and Parthasarathy*, 2009). Modelling and disentangle such large spatio-temporal relations across the globe is an open research problem. Third, it would be appealing to provide scores for drivers instead of a binary classification. This could be achieved by measuring the distance to the nearest code in the Lockup Free Quantization (LFQ). Forth, the prediction of the model depends on the capacity of reanalysis data to accurately represent the local environmental factors and land-atmospheric feedback. Most importantly, drawing conclusions on drivers from weak predictive models may lead to unreliable interpretations. Finally, our model does not identify causal relationships.

Despite these limitations, our approach demonstrates a clear capability in identifying drivers and anomalies in climate data which would allow a more timely event attribution during and right after extreme events.

8.2.4 Flood Forecasting

Extreme flood events, characterized by longer return periods, are expected to become more frequent and intense due to climate change. Traditional hydrology models often struggle to accurately predict such events. To improve flood detection, it is crucial to develop computationally efficient and precise deep learning models capable of forecasting key hydrological variables, such as river discharge. In Chapter 7, we demonstrated the potential of RiverMamba for predicting extreme riverine floods and

in Section 7.3.7, we presented case studies for extreme flood events. We see this as the primary motivation for developing RiverMamba. However, it is important to acknowledge that early warning systems may sometimes fail, leading to inaccurate forecasts. This limitation should be considered when deploying early warning systems. In addition, while the results show major advancements in river discharge and flood forecasting, the approach has some more limitations. For a real operational setting, only data can be used that is available until the current day t. For instance, ERA5-Land is publicly available after 5 days whereas we assumed that ERA5-Land is already available after 1 day, e.g., $t-1$. ERA5-Land, however, could be substituted by other near real-time reanalysis data that is earlier available or analysis data until day t. It also needs to be mentioned that observational data like GRDC (Section 2.2.7) are affected by human interventions like dams and there is a need to integrate such interventions in the model. As it is the case for operational systems, floods are not always correctly forecast. The causes of the errors need to be analyzed more in detail. The forecast errors can be caused by human interventions, errors in the weather forecast for meteorological forcing or river attributes, the rarity of floods, or bias in the data and re-analysis. Given such errors, it is desirable to extend the model such that it estimates its uncertainty for the forecast as well.

Besides these limitations, RiverMamba has the potential for an operational medium-range river discharge and flood forecasting system that predicts flood risks, in particular extreme floods, more accurately and at higher resolution than existing systems. This is essential for stakeholders to make decisions for an effective flood risk mitigation strategy and an early warning system to protect citizens.

8.3 Outlook and Future Works

8.3.1 Ensemble Forecast

Had we forecast multiple scenarios for a disaster earlier, we would have been able to prepare varied decisions that could lead to better adaptation strategies. Had we obtained the probabilities of various scenarios, we would have been able to quantify uncertainty and weigh the potential costs and benefits of different actions under different situations. These are examples of the importance of having a probabilistic forecast. All works presented in this dissertation are based on deterministic forecasts. That is, the models do not provide ensemble or probabilistic forecasts to account for different scenarios or to estimate uncertainty. In general, deterministic deep learning models for numerical weather prediction and flood forecasting still fall short in accuracy and reliability compared to their probabilistic counterparts (*Kratzert et al.*, 2019; *Price et al.*, 2025). This is because the observations of initial condition are incomplete and the forecast generated by weather models remain uncertain. Ensemble forecast mitigates the inherent sensitivity to initial conditions in a chaotic system. Current state-of-the-art models for probabilistic weather forecast use denoising diffusion (*Price et al.*, 2025) and flow matching (*Couairon et al.*, 2024) algorithms. In hydrology, denoising diffusion model was just recently explored for probabilistic flood forecasting but at a local scale (*Ou et al.*, 2025). However, denoising diffusion based generation is characterized with slow inference time compared to other generative algorithms and with a dense grid, the algorithm becomes even less efficient. Recently, *Alet et al.* (2025) proposed a new framework. The idea is to train a stochastic model ensemble (e.g., 4 models with injected noise in each to perturb the parameter space) to optimize the Continuous Ranked Probability Score (CRPS). In this way, the model became more efficient and skillful than

its diffusion counterpart models. Exploring scoring rule minimization in hydrology as in *Alet et al.* (2025) and *Lang et al.* (2024b) and adapting it on RiverMamba (Chapter 7) is a promising line of research which can help in generating maps of different flood scenarios (Fig. 8.1). Ensemble forecasts helps also dealing with the uncertainty in the river discharge values. As we have seen in Chapter 7, river discharge exhibits a very large dynamic range, meaning that the higher the discharge, the higher the uncertainty in the measurement and modeling.

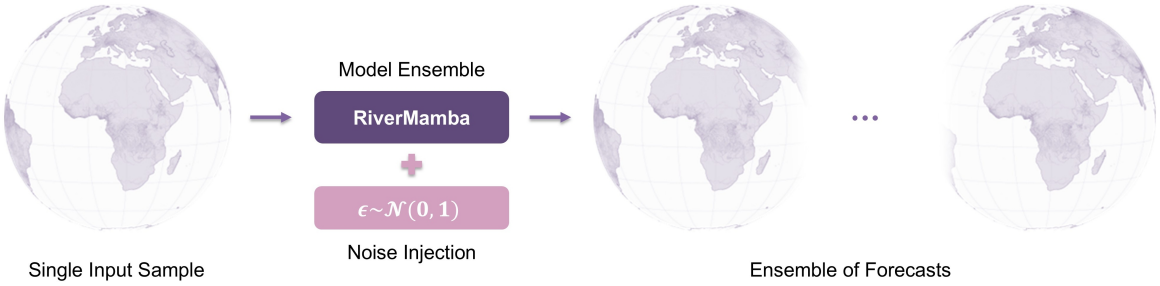


Figure 8.1: An overview of the probabilistic generative process of RiverMamba as in *Alet et al.* (2025).

8.3.2 Forecasting Vegetation Conditions from Climate Scenarios

Climate adaptation and agricultural decision-making depend on our ability to forecast drought conditions and vegetation health. In Chapter 5, we have shown how Focal-TSMP can predict vegetation health from a regional climate simulation. The proposed approach can be used to predict future trends in the vegetation dynamic based on climate scenarios. Providing this information, the model can help to recognize regions that are expected to be more vulnerable to agricultural drought risks. This is based on "what if" kind of experiments which are more beneficial for the general public and policymakers than for scientists. We have tested Focal-TSMP with a control run simulation or a baseline scenario (Section 2.2.2.1). Testing Focal-TSMP with a control run first was essential for the model development part. In addition, this control run simulation serves as a baseline for other scenarios. Yet, the robustness of the developed Focal-TSMP model under anthropogenic land- and water use changes remains underexplored. Focal-TSMP can be tested on a human water use scenario derived from TSMP and irrigation datasets like *Siebert and Döll* (2010) and *Wada et al.* (2016). It can be also tested with a land use land cover scenario (LULC) derived from TSMP and the European crop dynamics dataset from *Baumert et al.* (2025). In order to guarantee that vegetation condition variations result only from prescribed human interventions, the scenarios should be derived from the baseline scenario, ensuring persistent initial conditions and atmospheric forcing. Human interventions can be induced either by introducing additional forcings (adding or extracting water fluxes from the surface and subsurface in the ParFlow component) or by modifying the static input files (LULC files, e.g., by altering the crop composition over time while maintaining the overall fraction of cropland and grassland). Similarly, the developed model for identifying anomalies and the drivers of extremes impacts (Chapter 6) can be used for TSMP human water- and land use scenarios to quantify the impact of anthropogenic drivers on the distribution of anomalies in the state variables. For example, we can first train the model in Chapter 6 with a control run or reanalysis and then run the pretrained model on different scenarios. We can then analyze which scenario causes a statistically

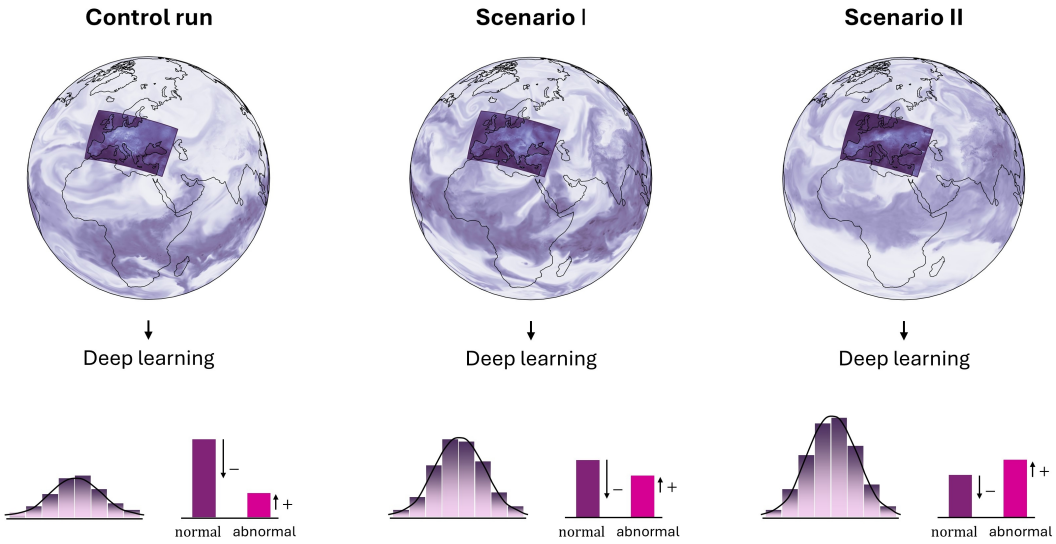


Figure 8.2: An illustration of testing the developed model in Chapter 6 for land- and water use changes scenarios.

significant change of the distribution of anomalies by comparing different scenarios with the control run (Fig. 8.2). This would enable us to identify potential drivers of anomalies such as land- and water use changes.

8.3.3 Agricultural Drought Forecasting

Although our Focal-TSMP model (Chapter 5) is trained to predict vegetation products as they would be observed from the AVHRR platform, it would be possible to predict target variables from different platforms or following different atmospheric corrections. This could be done as future work by training multiple deep learning models or by encoding different satellite systems similar to conditional image generation and train the model to distinguish between them. Moreover, our work can be extended to predict other vegetation products from different satellite platforms depending on requirements. The predicted satellite-based indices can be combined with different meteorological drought indices to provide more comprehensive drought assessments under future climate change.

We believe that our Focal-TSMP could be also useful to simulate remote sensing products from down-scaled simulations and to be used as a supportive evaluation framework to further investigate the predictive capability of the simulation to reproduce drought events and consequently to improve the TSMP model development. Finally, TSMP simulations still rely on prescribed vegetation condition, e.g., prescribed Leaf Area Index (LAI) from climatology as an input to the physics-based model. Focal-TSMP could be used to model LAI and thus provide TSMP simulation with dynamic LAI and more realistic vegetation responses under different scenarios.

8.3.4 Identifying Drivers of Extreme Events Impacts

Although we focused in Chapter 6 on vegetation indices as impacts of extreme events in our study, our approach is not limited to such extremes. Since we use a binary representation of extremes, our approach can also be applied to other extremes that cannot be derived from satellite products, but that

are stored in a binary format in databases.

In addition, in Chapter 6, we have shown the overall distribution of drivers and impacts of extremes, but we have not relate specific identified drivers to specific impacts. This can be done by using a pretrained model as in Chapter 6 with post-hoc explainable AI methods (e.g., integrated gradient) to isolate the drivers for selected regions or spots. For example, we can select a region where we have an extreme impact and only compute the gradient of this region with respect to the identified drivers after the quantization layer.

8.3.5 Flood Forecasting

RiverMamba presented in Chapter 7 can be seen as a general backbone model that is not limited to medium-range river discharge forecast and specific input variables or task. For instance, RiverMamba can be used for seasonal river discharge forecasting and land surface modeling. Furthermore, weather forecast and climate modeling generally require a large receptive field to preserve the dynamic over lead time and RiverMamba is a promising backbone for such task. In what follows, we briefly describe a potential research theme on the intersection of remote sensing, computer vision and hydrology. The overarching goal is to develop a deep learning model that operates directly from the observation space to forecast a probabilistic ensemble of accurate floods and key hydrological state variables.

8.3.5.1 Learning Directly from Observations

Thanks to the recent research efforts (*Bi et al.*, 2023; *Pathak et al.*, 2022; *Price et al.*, 2025), deep learning for numerical weather prediction has made it to the mainstream. These advances rely on reanalysis data like ERA5 (*Hersbach et al.*, 2020) to initiate the forecast. Reanalysis is generated via data assimilation, a framework that combines physics-based simulation with Earth observations to generate a consistent current state of key land and atmospheric variables. But depending on reanalysis inherits the biases from the data assimilation process. In addition, reanalysis is not available in real time which could be crucial in operational setting. That is why recent works like *McNally et al.* (2024) and *Alexe et al.* (2024) started exploring to learn directly from the observation space. This has numerous advantages, including computational efficiency and overcoming the information loss. In hydrology, state-of-the-art deep learning models (*Nearing et al.*, 2024) for river discharge forecasting still use ERA5 as initial condition. In our RiverMamba work (Chapter 7), we also used ERA5-Land as an initial condition. One solution is to substitute ERA5-Land by Near-Real-Time ERA5-Land/ERA5 or by analysis data. Another interesting line of research can be pursued to replace reanalysis by near real-time multi-source remote sensing satellite observations. This comes surely with its own challenges since observations are irregular in space and time.

8.3.5.2 Adaptive Forecasts with Multiple Weather Models

As we have seen in Section 7.3.8.5, Meteorological forcing forecasts, e.g., precipitation forecast has a big impact on flood forecasting, but which weather model is the best? Weather models perform differently around the globe. In our RiverMamba work (Chapter 7), we used the ECMWF weather forecast similar to GloFAS (*Harrigan et al.*, 2023). The work can be improved by additionally incorporating weather models like Global Ensemble Forecast System (GEFS) and ICON (*Zängl et al.*,

2015). The aim is to design an adaptive model for flood forecasting that incorporates information from different weather forecasting products to optimize the forecast with lead time. Here forecast derived from machine learning models like AIFS (*Lang et al.*, 2024a) can be tested as well. Google's flood forecasting initiative has taken a step in this direction (*Shalev and Kratzert*, 2024).

8.3.5.3 High-resolution River Discharge over Europe

The European Flood Awareness System (EFAS) provides consistent representation of hydrological variables across Europe with high spatial resolution (0.017° compared to 0.05° in GloFAS) and sub-daily temporal resolution from nowcasting to seasonal forecast (Fig. 8.3).

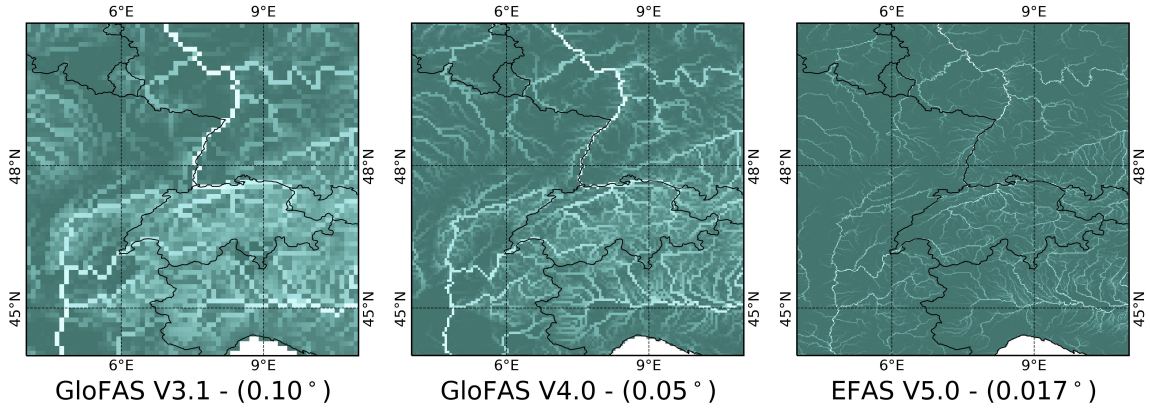


Figure 8.3: An overview of the same product output generated by different hydrological systems.

The EFAS research project was started long time ago ($\sim$ 2003 (*Thielen et al.*, 2009)) and impressive progresses have been made in developing and improving EFAS. However, operating EFAS for real-time and time-critical operations is still computationally expensive. The forecast become available for the general public with a delay of many days. Starting from RiverMamba (Chapter 7), we believe that the model has the potential to be employed on other dataset with a higher resolution. A promising research direction is thus to use the techniques used in RiverMamba for training a deep learning model on EFAS data. This could play a critical role in helping stakeholders make informed decisions for better flood risk mitigation and early warning strategies in Europe. Here adapting the ICON model for weather forecast (*Zängl et al.*, 2015) can be tested along with the ECMWF weather forecast.

8.3.5.4 Down-scaling the Global Flood Awareness System (GloFAS)

Hydrological systems like EFAS are expensive to run. This is because the resolution is high in space and time. EFAS covers only the European domain, thus people around the world use GloFAS to access reanalysis and forecast products. There exists advanced deep learning approaches for atmospheric down-scaling (*Mardani et al.*, 2025). Similar methodologies can be applied with RiverMamba (Chapter 7) for down-scaling from GloFAS to EFAS. While the short-term goal is to run EFAS more efficiently, the long-term goal is to apply the pre-trained model on other regions of the Earth where high-resolution products do not exist, e.g., by using LOAN (Chapter 4) for conditioning the image generation on low-resolution GloFAS, the river networks and other static river attributes.

Location Embeddings can be used as a conditioning as well (*Klemmer et al.*, 2025; *Brown et al.*, 2025).

8.3.5.5 Sub-seasonal to Seasonal (S2S) River Discharge Forecasting

In our RiverMamba work (Chapter 7), we have shown that RiverMamba can forecast medium-range floods and river discharge. Although short- and medium-range forecasting, e.g., for a few days, is valuable for immediate planning, more benefits would be realized with longer forecasting time. Building on the methodology presented in Chapter 7 but with different data, a deep learning model for sub-seasonal to seasonal (S2S) river discharge and flood forecasting can be achieved.

Code and Data Availability

- Code and data for Chapter 4:
  - Code: The source code and pretrained models to reproduce the results are available at https://github.com/HakamShams/LOAN.
  - Dataset: The training and testing samples from FireCube dataset can be downloaded from https://zenodo.org/record/6528394 (*Prapas et al.*, 2022c).
- Code and data for Chapter 5:
  - Code: The source code and pretrained models to reproduce the results are published at https://zenodo.org/records/10015049 (*Shams Eddin and Gall*, 2023a). The source code is also available on GitHub at https://github.com/HakamShams/Focal_TSMP.
  - Dataset: The pre-processed data used in the study are available at https://doi.org/10.5281/zenodo.10008815 (*Shams Eddin and Gall*, 2023b). The original TSMP data are stored at Jülich Research Centre at https://www.re3data.org/repository/r3d100012923 (*Furusho-Percot et al.*, 2019a), as well as at PANGAEA at https://doi.pangaea.de/10.1594/PANGAEA.901823 (*Furusho-Percot et al.*, 2019b). The raw vegetation health products can be downloaded from the National Oceanic and Atmospheric Administration (NOAA/NESDIS), Center for Satellite Applications and Research (STAR) at https://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/index.php (*Yang et al.*, 2020).
- Code and data for Chapter 6:
  - Code: The source code and pretrained models to reproduce the results are available on GitHub at https://github.com/HakamShams/IDEE. The source code for the synthetic data generation is also available on GitHub at https://github.com/HakamShams/Synthetic_Multivariate_Anomalies.
  - Dataset: The pre-processed data used in the study are available at https://doi.org/10.60507/FK2/RD9E33 (*Shams Eddin and Gall*, 2024b). The raw vegetation health products can be downloaded from the National Oceanic and Atmospheric Administration (NOAA/NESDIS), Center for Satellite Applications and Research (STAR) at https://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/index.php (*Yang et al.*, 2020). The raw CERRA sub-daily regional reanalysis data for Europe on single levels can be downloaded from Copernicus Climate Change Service (C3S), Climate Data Store (CDS) at https://doi.org/10.24381/cds.622a565a (*Schimanke et al.*, 2021). The raw ERA5-Land hourly reanalysis data can be downloaded from Copernicus Climate Change Service (C3S), Climate Data Store (CDS) at https://doi.org/10.24381/cds.e2161bac (*Muñoz Sabater et al.*, 2021).

Code and data for Chapter 7:

- Code: The code of RiverMamba, processing scripts, and pretrained models are available on GitHub at https://github.com/HakamShams/RiverMamba_code.
- Dataset: The pre-processed data used in the study and RiverMamba reforecasts are available at https://doi.org/10.60507/FK2/T8QYWE (*Shams Eddin et al.*, 2025b). GRDC data that has been used in the study is available for researchers after signing a license agreement with the owner of the data at https://grdc.bafg.de/. Instructions on how the data can be obtained and used are provided in the source code. Raw ECMWF-IFS HRES weather forecasts can be downloaded from the ECMWF archive catalogue (MARS). The raw River discharge and related historical data from the Global Flood Awareness System (GloFAS) can be downloaded from the Early Warning Data Store (EWDS) at https://doi.org/10.24381/cds.a4fdd6b9 (*Harrigan et al.*, 2020). The raw LISFLOOD static and parameter maps can be downloaded from the European Commission, Joint Research Centre (JRC) at http://data.europa.eu/89h/68050d73-9c06-499c-a441-dc5053cb0c86 (*Choulga et al.*, 2024). The raw HydroRIVERS and HydroATLAS static data can be downloaded from https://www.hydrosheds.org/products/hydrorivers (*Lehner and Grill*, 2013; *Linke et al.*, 2019). The raw CPC data can be downloaded from the National Oceanic and Atmospheric Administration (NOAA), Climate Prediction Center (CPC) at https://psl.noaa.gov/data/gridded/data.cpc.globalprecip.html. The raw ERA5-Land hourly reanalysis data can be downloaded from Copernicus Climate Change Service (C3S), Climate Data Store (CDS) at https://doi.org/10.24381/cds.e2161bac (*Muñoz Sabater et al.*, 2021).

Bibliography

Acuña Espinoza, E.; Kratzert, F.; Klotz, D.; Gauch, M.; Álvarez Chaves, M.; Loritz, R., and Ehret, U. Technical note: An approach for handling multiple temporal frequencies with different input dimensions using a single lstm cell. *Hydrology and Earth System Sciences*, 29(6):1749–1758, 2025. doi: 10.5194/hess-29-1749-2025. URL https://hess.copernicus.org/articles/29/1749/2025/. (Cited on page 47.)

Adede, Chrisgone; Oboko, Robert; Wagacha, Peter Waiganjo, and Atzberger, Clement. A mixed model approach to vegetation condition prediction using artificial neural networks (ann): Case of kenya's operational drought monitoring. *Remote Sensing*, 11(9), 2019. ISSN 2072-4292. doi: 10.3390/rs11091099. URL https://www.mdpi.com/2072-4292/11/9/1099. (Cited on page 46.)

Aleissaee, Abdulaziz Amer; Kumar, Amandeep; Anwer, Rao Muhammad; Khan, Salman; Cholakkal, Hisham; Xia, Gui-Song, and Khan, Fahad Shahbaz. Transformers in remote sensing: A survey. *Remote Sensing*, 15(7), 2023. ISSN 2072-4292. doi: 10.3390/rs15071860. URL https://www.mdpi.com/2072-4292/15/7/1860. (Cited on page 88.)

Alet, Ferran; Price, Ilan; El-Kadi, Andrew; Masters, Dominic; Markou, Stratis; Andersson, Tom R; Stott, Jacklynn; Lam, Remi; Willson, Matthew; Sanchez-Gonzalez, Alvaro, and others, . Skillful joint probabilistic weather forecasting from marginals. *arXiv preprint arXiv:2506.10772*, 2025. (Cited on pages 215 and 216.)

Alexe, Mihai; Boucher, Eulalie; Lean, Peter; Pinnington, Ewan; Laloyaux, Patrick; McNally, Anthony; Lang, Simon; Chantry, Matthew; Burrows, Chris; Chrust, Marcin, and others, . Graphdop: Towards skilful data-driven medium-range weather forecasts learnt and initialised directly from observations. *arXiv preprint arXiv:2412.15687*, 2024. (Cited on page 218.)

Alfieri, L.; Burek, P.; Dutra, E.; Krzeminski, B.; Muraro, D.; Thielen, J., and Pappenberger, F. Glofas - global ensemble streamflow forecasting and flood early warning. *Hydrology and Earth System Sciences*, 17(3):1161–1175, 2013. doi: 10.5194/hess-17-1161-2013. URL https://hess.copernicus.org/articles/17/1161/2013/. (Cited on pages 30 and 155.)

Amari, S.-I. Learning patterns and pattern sequences by self-organizing nets of threshold elements. *IEEE Transactions on Computers*, C-21(11):1197–1206, 1972. doi: 10.1109/T-C.1972.223477. (Cited on page 12.)

Astrid, Marcella; Zaheer, Muhammad Zaigham, and Lee, Seung-Ik. Synthetic temporal anomaly guided end-to-end video anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 207–214, 2021. (Cited on pages 44, 117, 118, 128, 130 and 131.)

Ba, Jimmy Lei; Kiros, Jamie Ryan, and Hinton, Geoffrey E. Layer normalization. *NIPS Deep Learning Symposium*, 2016. (Cited on pages 16, 48 and 83.)

Baldauf, Michael; Seifert, Axel; Förstner, Jochen; Majewski, Detlev; Raschendorfer, Matthias, and Reinhardt, Thorsten. Operational convective-scale numerical weather prediction with the cosmo model: Description and sensitivities. *Monthly Weather Review*, 139(12):3887 – 3905, 2011. doi: 10.1175/MWR-D-10-05013.1. URL https://journals.ametsoc.org/view/journals/mwre/139/12/mwr-d-10-05013.1.xml. (Cited on page 21.)

Balsamo, Gianpaolo; Beljaars, Anton; Scipal, Klaus; Viterbo, Pedro; van den Hurk, Bart; Hirschi, Martin, and Betts, Alan K. A revised hydrology for the ecmwf model: Verification from field site to terrestrial water storage and impact in the integrated forecast system. *Journal of Hydrometeorology*, 10(3):623 – 643, 2009. doi: 10.1175/2008JHM1068.1. URL https://journals.ametsoc.org/view/journals/hydr/10/3/2008jhm1068_1.xml. (Cited on page 30.)

Barcikowska, M. J.; Kapnick, S. B.; Krishnamurty, L.; Russo, S.; Cherchi, A., and Folland, C. K. Changes in the future summer mediterranean climate: contribution of teleconnections and local factors. *Earth System Dynamics*, 11(1):161–181, 2020. doi: 10.5194/esd-11-161-2020. URL https://esd.copernicus.org/articles/11/161/2020/. (Cited on page 52.)

Bashfield, A and Keim, A. Continent-wide dem creation for the european union. In *34th International Symposium on Remote Sensing of Environment. The GEOSS Era: Towards Operational Environmental Monitoring. Sydney, Australia*, pages 10–15, 2011. (Cited on page 20.)

Baumert, Josef; Heckelei, Thomas, and Storm, Hugo. A dataset of yearly probabilistic crop type maps for the eu from 1990 to 2018. *Data in Brief*, 60:111472, 2025. ISSN 2352-3409. doi: 10.1016/j.dib.2025.111472. URL https://www.sciencedirect.com/science/article/pii/S2352340925002045. (Cited on page 216.)

Baur, F.; Scheck, L.; Stumpf, C.; Köpken-Watts, C., and Potthast, R. A neural network-based method for generating synthetic 1.6 μm near-infrared satellite images. *EGUsphere*, 2023:1–31, 2023. doi: 10.5194/egusphere-2023-353. URL https://egusphere.copernicus.org/preprints/2023/egusphere-2023-353/. (Cited on page 104.)

Beck, Maximilian; Pöppel, Korbinian; Spanring, Markus; Auer, Andreas; Prudnikova, Oleksandra; Kopp, Michael; Klambauer, Günter; Brandstetter, Johannes, and Hochreiter, Sepp. xlstm: Extended long short-term memory. *Advances in Neural Information Processing Systems*, 37:107547–107603, 2024. (Cited on page 15.)

Bengio, Y.; Simard, P., and Frasconi, P. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2):157–166, 1994. doi: 10.1109/72.279181. (Cited on page 14.)

Benson, Vitus; Robin, Claire; Requena-Mesa, Christian; Alonso, Lazaro; Carvalhais, Nuno; Cortés, José; Gao, Zhihan; Linscheid, Nora; Weynants, Mélanie, and Reichstein, Markus. Multi-modal learning for geospatial vegetation forecasting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 27788–27799, June 2024. (Cited on page 46.)

Bentivoglio, R.; Isufi, E.; Jonkman, S. N., and Taormina, R. Deep learning methods for flood mapping: a review of existing applications and future research directions. *Hydrology and Earth System Sciences*, 26(16):4345–4378, 2022. doi: 10.5194/hess-26-4345-2022. URL https://hess.copernicus.org/articles/26/4345/2022/. (Cited on page 47.)

Bergado, John Ray; Persello, Claudio; Reinke, Karin, and Stein, Alfred. Predicting wildfire burns from big geodata using deep learning. *Safety science*, 140:105276, 2021. (Cited on page 45.)

Bergen, Karianne J.; Johnson, Paul A.; de Hoop, Maarten V., and Beroza, Gregory C. Machine learning for data-driven discovery in solid earth geoscience. *Science*, 363(6433):eaau0323, 2019. doi: 10.1126/science.aau0323. URL https://www.science.org/doi/abs/10.1126/science.aau0323. (Cited on page 46.)

Bergmann, Paul; Fauser, Michael; Sattlegger, David, and Steger, Carsten. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (CVPR)*, pages 4183–4192, 2020. (Cited on page 44.)

Bertasius, Gedas; Wang, Heng, and Torresani, Lorenzo. Is space-time attention all you need for video understanding? In *Proceedings of the International Conference on Machine Learning (ICML)*, July 2021. (Cited on pages 17, 60, 61, 62, 63 and 69.)

Betancourt, C.; Stomberg, T. T.; Edrich, A.-K.; Patnala, A.; Schultz, M. G.; Roscher, R.; Kowalski, J., and Stadtler, S. Global, high-resolution mapping of tropospheric ozone – explainable machine learning and impact of uncertainties. *Geoscientific Model Development*, 15(11):4331–4354, 2022. doi: 10.5194/gmd-15-4331-2022. URL https://gmd.copernicus.org/articles/15/4331/2022/. (Cited on page 104.)

Bi, Kaifeng; Xie, Lingxi; Zhang, Hengheng; Chen, Xin; Gu, Xiaotao, and Tian, Qi. Accurate medium-range global weather forecasting with 3d neural networks. *Nature*, 619(7970):533–538, 2023. doi: https://doi.org/10.1038/s41586-023-06185-3. (Cited on pages 155 and 218.)

Bindas, Tadd; Tsai, Wen-Ping; Liu, Jiangtao; Rahmani, Farshid; Feng, Dapeng; Bian, Yuchen; Lawson, Kathryn, and Shen, Chaopeng. Improving river routing using a differentiable muskingum-cunge model and physics-informed machine learning. *Water Resources Research*, 60(1):e2023WR035337, 2024. doi: https://doi.org/10.1029/2023WR035337. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023WR035337. e2023WR035337 2023WR035337. (Cited on page 48.)

Bjånes, Alexandra; De La Fuente, Rodrigo, and Mena, Pablo. A deep learning ensemble model for wildfire susceptibility mapping. *Ecological Informatics*, 65:101397, 2021. ISSN 1574-9541. doi: https://doi.org/10.1016/j.ecoinf.2021.101397. URL https://www.sciencedirect.com/science/article/pii/S1574954121001886. (Cited on page 45.)

Blöschl, Günter; Hall, Julia; Parajka, Juraj; Perdigão, Rui A. P.; Merz, Bruno; Arheimer, Berit; Aronica, Giuseppe T.; Bilibashi, Ardian; Bonacci, Ognjen; Borga, Marco; Čanjevac, Ivan; Castellarin, Attilio; Chirico, Giovanni B.; Claps, Pierluigi; Fiala, Károly; Frolova, Natalia; Gorbachova, Liudmyla; Gül, Ali; Hannaford, Jamie; Harrigan, Shaun; Kireeva, Maria; Kiss, An-

drea; Kjeldsen, Thomas R.; Kohnová, Silvia; Koskela, Jarkko J.; Ledvinka, Ondrej; Macdonald, Neil; Mavrova-Guirguinova, Maria; Mediero, Luis; Merz, Ralf; Molnar, Peter; Montanari, Alberto; Murphy, Conor; Osuch, Marzena; Ovcharuk, Valeryia; Radevski, Ivan; Rogger, Magdalena; Salinas, José L.; Sauquet, Eric; Šraj, Mojca; Szolgay, Jan; Viglione, Alberto; Volpi, Elena; Wilson, Donna; Zaimi, Klodian, and Živković, Nenad. Changing climate shifts timing of european floods. *Science*, 357(6351):588–590, 2017. doi: 10.1126/science.aan2506. URL https://www.science.org/doi/abs/10.1126/science.aan2506. (Cited on page 1.)

Blöschl, Günter; Hall, Julia; Viglione, Alberto; Perdigão, Rui AP; Parajka, Juraj; Merz, Bruno; Lun, David; Arheimer, Berit; Aronica, Giuseppe T; Bilibashi, Ardian, and others, . Changing climate both increases and decreases european river floods. *Nature*, 573(7772):108–111, 2019. doi: 10.1038/s41586-019-1495-6. (Cited on page 1.)

Bomers, Anouk and Hulscher, Suzanne J. M. H. Neural networks for fast fluvial flood predictions: Too good to be true? *River Research and Applications*, 39(8):1652–1658, 2023. doi: https://doi.org/10.1002/rra.4144. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/rra.4144. (Cited on page 47.)

Bountos, Nikolaos Ioannis; Sdraka, Maria; Zavras, Angelos; Karasante, Ilektra; Karavias, Andreas; Herekakis, Themistocles; Thanasou, Angeliki; Michail, Dimitrios, and Papoutsis, Ioannis. Kuro siwo: 33 billion m^2 under the water. a global multi-temporal satellite dataset for rapid flood mapping. In *Advances in Neural Information Processing Systems*, volume 37, pages 38105–38121. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/43612b0662cb6a4986edf859fd6ebafe-Paper-Datasets_and_Benchmarks_Track.pdf. (Cited on page 47.)

Boussard, Julien; Nagda, Chandni; Kaltenborn, Julia; Lange, Charlotte Emilie Elektra; Brouillard, Philippe; Gurwicz, Yaniv; Nowack, Peer, and Rolnick, David. Towards causal representations of climate model data. *arXiv preprint arXiv:2312.02858*, 2023. (Cited on page 3.)

Brehmer, Johann; De Haan, Pim; Lippe, Phillip, and Cohen, Taco S. Weakly supervised causal representation learning. *Advances in Neural Information Processing Systems*, 35:38319–38331, 2022. (Cited on page 3.)

Breiman, Leo. Random forests. *Machine Learning 2001 45:1*, 45:5–32, 10 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL https://link.springer.com/article/10.1023/A:1010933404324. (Cited on page 59.)

Brown, Christopher F; Kazmierski, Michal R; Pasquarella, Valerie J; Rucklidge, William J; Samsikova, Masha; Zhang, Chenhui; Shelhamer, Evan; Lahera, Estefania; Wiles, Olivia; Ilyushchenko, Simon, and others, . Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. *arXiv preprint arXiv:2507.22291*, 2025. (Cited on page 220.)

Burrichter, Benjamin; Hofmann, Julian; Koltermann da Silva, Juliana; Niemann, Andre, and Quirmbach, Markus. A spatiotemporal deep learning approach for urban pluvial flood forecasting with

multi-source data. *Water*, 15(9), 2023. ISSN 2073-4441. doi: 10.3390/w15091760. URL https://www.mdpi.com/2073-4441/15/9/1760. (Cited on page 47.)

Busker, Tim; van den Hurk, Bart; de Moel, Hans, and Aerts, Jeroen CJH. The value of precipitation forecasts to anticipate floods. *Bulletin of the American Meteorological Society*, 2025. (Cited on page 47.)

Büttner, György. *CORINE Land Cover and Land Cover Change Products*, pages 55–74. Springer Netherlands, Dordrecht, 2014. ISBN 978-94-007-7969-3. doi: 10.1007/978-94-007-7969-3_5. URL https://doi.org/10.1007/978-94-007-7969-3_5. (Cited on page 20.)

Cache, T.; Gomez, M. S.; Beucler, T.; Blagojevic, J.; Leitao, J. P., and Peleg, N. Enhancing generalizability of data-driven urban flood models by incorporating contextual information. *Hydrology and Earth System Sciences*, 28(24):5443–5458, 2024. doi: 10.5194/hess-28-5443-2024. URL https://hess.copernicus.org/articles/28/5443/2024/. (Cited on page 47.)

Cammalleri, Carmelo; Vogt, Jürgen V; Bisselink, Bernard, and de Roo, Ad. Comparing soil moisture anomalies from multiple independent sources over different regions across the globe. *Hydrology and Earth System Sciences*, 21(12):6329–6343, 2017. (Cited on page 20.)

Camps-Valls, Gustau; Fernández-Torres, Miguel-Ángel; Cohrs, Kai-Hendrik; Höhl, Adrian; Castelletti, Andrea; Pacal, Aytac; Robin, Claire; Martinuzzi, Francesco; Papoutsis, Ioannis; Prapas, Ioannis, and others, . Artificial intelligence for modeling and understanding extreme weather and climate events. *Nature Communications*, 16(1):1919, 2025. doi: 10.1038/s41467-025-56573-8. (Cited on pages 2 and 155.)

Cao, Shengpeng; Zhang, Lifeng; He, Yi; Zhang, Yali; Chen, Yi; Yao, Sheng; Yang, Wang, and Sun, Qiang. Effects and contributions of meteorological drought on agricultural drought under different climatic zones and vegetation types in northwest china. *Science of The Total Environment*, 821:153270, 2022. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2022.153270. URL https://www.sciencedirect.com/science/article/pii/S0048969722003618. (Cited on pages 43 and 108.)

Carreira, João and Zisserman, Andrew. Quo vadis, action recognition? a new model and the kinetics dataset. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4724–4733, 2017. doi: 10.1109/CVPR.2017.502. (Cited on page 14.)

Chen, Guo; Huang, Yifei; Xu, Jilan; Pei, Baoqi; Chen, Zhe; Li, Zhiqi; Wang, Jiahao; Li, Kunchang; Lu, Tong, and Wang, Limin. Video mamba suite: State space model as a versatile alternative for video understanding. *arXiv preprint arXiv:2403.09626*, 2024. (Cited on page 50.)

Chen, Mingyue; Shi, Wei; Xie, Pingping; Silva, Viviane BS; Kousky, Vernon E; Wayne Higgins, R, and Janowiak, John E. Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research: Atmospheres*, 113(D4), 2008. (Cited on page 35.)

Chen, Tianqi and Guestrin, Carlos. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA, 2016. Association for Computing Machin-

ery. ISBN 9781450342322. doi: 10.1145/2939672.2939785. URL https://doi.org/10.1145/2939672.2939785. (Cited on page 59.)

Chen, Ting; Lucic, Mario; Houlsby, Neil, and Gelly, Sylvain. On self modulation for generative adversarial networks. In *International Conference on Learning Representations (ICLR)*, 2019. URL https://openreview.net/forum?id=Hkl5aoR5tm. (Cited on page 48.)

Chen, Yingxian; Liu, Zhengzhe; Zhang, Baoheng; Fok, Wilton; Qi, Xiaojuan, and Wu, Yik-Chung. Mgfn: Magnitude-contrastive glance-and-focus network for weakly-supervised video anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 387–395, 2023. (Cited on page 45.)

Chen, Zhiting; Liu, Hongyan; Xu, Chongyang; Wu, Xiuchen; Liang, Boyi; Cao, Jing, and Chen, Deliang. Modeling vegetation greenness and its climate sensitivity with deep-learning technology. *Ecology and Evolution*, 11(12):7335–7345, 2021. doi: 10.1002/ece3.7564. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.7564. (Cited on pages 46 and 47.)

Chiu, Chung-Cheng; Qin, James; Zhang, Yu; Yu, Jiahui, and Wu, Yonghui. Self-supervised learning with random-projection quantizer for speech recognition. In *International Conference on Machine Learning*, pages 3915–3924. PMLR, 2022. (Cited on pages 137 and 138.)

Choulga, M.; Moschini, F.; Mazzetti, C.; Grimaldi, S.; Disperati, J.; Beck, H.; Salamon, P., and Prudhomme, C. Technical note: Surface fields for global environmental modelling. *Hydrology and Earth System Sciences*, 28(13):2991–3036, 2024. doi: 10.5194/hess-28-2991-2024. URL https://hess.copernicus.org/articles/28/2991/2024/. (Cited on pages 35 and 224.)

Christian, Jordan I; Basara, Jeffrey B; Hunt, Eric D; Otkin, Jason A; Furtado, Jason C; Mishra, Vimal; Xiao, Xiangming, and Randall, Robb M. Global distribution, trends, and drivers of flash drought occurrence. *Nature communications*, 12(1):6330, 2021. doi: 10.1038/s41467-021-26692-z. (Cited on page 79.)

Christian, Jordan I; Martin, Elinor R; Basara, Jeffrey B; Furtado, Jason C; Otkin, Jason A; Lowman, Lauren EL; Hunt, Eric D; Mishra, Vimal, and Xiao, Xiangming. Global projections of flash drought show increased risk in a warming climate. *Communications Earth & Environment*, 4(1): 165, 2023. doi: 10.1038/s43247-023-00826-1. (Cited on pages 1 and 79.)

Chu, Haibo; Wu, Wenyan; Wang, Q.J.; Nathan, Rory, and Wei, Jiahua. An ann-based emulation modelling framework for flood inundation modelling: Application, challenges and future directions. *Environmental Modelling & Software*, 124:104587, 2020. ISSN 1364-8152. doi: https://doi.org/10.1016/j.envsoft.2019.104587. URL https://www.sciencedirect.com/science/article/pii/S1364815219307601. (Cited on page 47.)

Collin, Anne-Sophie and De Vleeschouwer, Christophe. Improved anomaly detection by training an autoencoder with skip connections on images corrupted with stain-shaped noise. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 7915–7922. IEEE, 2021. (Cited on page 44.)

Coogan, Sean C. P.; Robinne, François; Jain, Piyush, and Flannigan, Mike D. Scientists' warning on wildfire - a canadian perspective. *Canadian Journal of Forest Research*, 2019. (Cited on page 52.)

Couairon, Guillaume; Singh, Renu; Charantonis, Anastase; Lessig, Christian, and Monteleoni, Claire. Archesweather & archesweathergen: a deterministic and generative model for efficient ml weather forecasting. *arXiv preprint arXiv:2412.12971*, 2024. (Cited on page 215.)

Dahoui, Mohamed. Use of machine learning for the detection and classification of observation anomalies, 01/2023 2023. (Cited on page 43.)

Dai, Meng; Huang, Shengzhi; Huang, Qiang; Zheng, Xudong; Su, Xiaoling; Leng, Guoyong; Li, Ziyan; Guo, Yi; Fang, Wei, and Liu, Yongjia. Propagation characteristics and mechanism from meteorological to agricultural drought in various seasons. *Journal of Hydrology*, 610:127897, 2022. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2022.127897. URL https://www.sciencedirect.com/science/article/pii/S0022169422004723. (Cited on page 108.)

Dao, Tri. FlashAttention-2: Faster attention with better parallelism and work partitioning. In *International Conference on Learning Representations (ICLR)*, 2024. (Cited on pages 17 and 204.)

Dao, Tri and Gu, Albert. Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. In *International Conference on Machine Learning (ICML)*, 2024. (Cited on page 205.)

Dao, Tri; Fu, Dan; Ermon, Stefano; Rudra, Atri, and Ré, Christopher. Flashattention: Fast and memory-efficient exact attention with io-awareness. *Advances in neural information processing systems*, 35:16344–16359, 2022. (Cited on pages 17 and 204.)

Das, Mahashweta and Parthasarathy, Srinivasan. Anomaly detection and spatio-temporal analysis of global climate system. In *Proceedings of the Third International Workshop on Knowledge Discovery from Sensor Data*, SensorKDD '09, page 142–150, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605586687. doi: 10.1145/1601966.1601989. URL https://doi.org/10.1145/1601966.1601989. (Cited on pages 43 and 214.)

Das, Monidipa and Ghosh, Soumya K. Deep-step: A deep learning approach for spatiotemporal prediction of remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 13(12):1984–1988, 2016. doi: 10.1109/LGRS.2016.2619984. (Cited on page 46.)

Dasgupta, Antara; Hostache, Renaud; Ramsankaran, RAAJ; Schumann, Guy J.-P.; Grimaldi, Stefania; Pauwels, Valentijn R. N., and Walker, Jeffrey P. A mutual information-based likelihood function for particle filter flood extent assimilation. *Water Resources Research*, 57 (2):e2020WR027859, 2021. doi: https://doi.org/10.1029/2020WR027859. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020WR027859. e2020WR027859 2020WR027859. (Cited on page 47.)

de Burgh-Day, C. O. and Leeuwenburg, T. Machine learning for numerical weather and climate modelling: a review. *Geoscientific Model Development*, 16(22):6433–6477, 2023. doi: 10.5194/gmd-16-6433-2023. URL https://gmd.copernicus.org/articles/16/6433/2023/. (Cited on page 46.)

Dee, D. P.; Uppala, S. M.; Simmons, A. J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M. A.; Balsamo, G.; Bauer, P.; Bechtold, P.; Beljaars, A. C. M.; van de Berg, L.; Bidlot, J.; Bormann, N.; Delsol, C.; Dragani, R.; Fuentes, M.; Geer, A. J.; Haimberger, L.; Healy, S. B.; Hersbach, H.; Hólm, E. V.; Isaksen, L.; Kållberg, P.; Köhler, M.; Matricardi, M.; McNally, A. P.; Monge-Sanz, B. M.; Morcrette, J.-J.; Park, B.-K.; Peubey, C.; de Rosnay, P.; Tavolato, C.; Thépaut, J.-N., and Vitart, F. The era-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597, 2011. doi: 10.1002/qj.828. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.828. (Cited on page 21.)

Defard, Thomas; Setkov, Aleksandr; Loesch, Angelique, and Audigier, Romaric. Padim: A patch distribution modeling framework for anomaly detection and localization. In *Pattern Recognition. ICPR International Workshops and Challenges*, pages 475–489, Cham, 2021. Springer International Publishing. ISBN 978-3-030-68799-1. (Cited on page 44.)

Deng, Hanqiu and Li, Xingyu. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9737–9746, June 2022. (Cited on page 44.)

Deng, Jia; Dong, Wei; Socher, Richard; Li, Li-Jia; Li, Kai, and Fei-Fei, Li. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 248–255, 2009. doi: 10.1109/CVPR.2009.5206848. (Cited on pages 87 and 88.)

Deng, Ziqing; Wang, Yang; Zhang, Bing; Li, Linwei; Wang, Jihong; Bian, Lifeng, and Yang, Chen. A triple-path spectral–spatial network with interleave-attention for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15: 5906–5923, 2022. doi: 10.1109/JSTARS.2022.3192470. (Cited on page 49.)

Di Giuseppe, Francesca; Vitolo, Claudia; Barnard, Christopher; Libertá, Giorgio; Maciel, Pedro; San-Miguel-Ayanz, Jesus; Villaume, Sebastien, and Wetterhall, Fredrik. Global seasonal prediction of fire danger. *Scientific data*, 11(1):128, 2024. doi: 10.1038/s41597-024-02948-3. (Cited on page 45.)

Diaconu, Codruţ-Andrei; Saha, Sudipan; Günnemann, Stephan, and Xiang Zhu, Xiao. Understanding the role of weather data for earth surface forecasting using a convlstm-based model. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1361–1370, 2022. doi: 10.1109/CVPRW56347.2022.00142. (Cited on page 46.)

Didan, Kamel. Mod13a2 modis/terra vegetation indices 16-day l3 global 1km sin grid v006 [data set]. nasa eosdis land processes daac, 2015. URL https://doi.org/10.5067/MODIS/MOD13A2.006. (Cited on page 20.)

Dosovitskiy, Alexey; Beyer, Lucas; Kolesnikov, Alexander; Weissenborn, Dirk; Zhai, Xiaohua; Unterthiner, Thomas; Dehghani, Mostafa; Minderer, Matthias; Heigold, Georg; Gelly, Sylvain; Uszkoreit, Jakob, and Houlsby, Neil. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy. (Cited on page 80.)

Dottori, F.; Kalas, M.; Salamon, P.; Bianchi, A.; Alfieri, L., and Feyen, L. An operational procedure for rapid flood risk assessment in europe. *Natural Hazards and Earth System Sciences*, 17(7): 1111–1126, 2017. doi: 10.5194/nhess-17-1111-2017. URL https://nhess.copernicus.org/articles/17/1111/2017/. (Cited on pages 31 and 155.)

Dottori, Francesco; Szewczyk, Wojciech; Ciscar, Juan-Carlos; Zhao, Fang; Alfieri, Lorenzo; Hirabayashi, Yukiko; Bianchi, Alessandra; Mongelli, Ignazio; Frieler, Katja; Betts, Richard A, and others, . Increased human and economic losses from river flooding with anthropogenic warming. *Nature Climate Change*, 8(9):781–786, 2018. doi: https://doi.org/10.1038/s41558-018-0257-z. (Cited on page 154.)

Drees, Lukas; Weber, Immanuel; Rußwurm, Marc, and Roscher, Ribana. Time dependent image generation of plants from incomplete sequences with cnn-transformer. In *DAGM German Conference on Pattern Recognition*, pages 495–510. Springer, 2022. (Cited on page 50.)

Düben, Peter; Modigliani, Umberto; Geer, Alan; Siemen, Stephan; Pappenberger, Florian; Bauer, Peter; Brown, Andy; Palkovic, Martin; Raoult, Baudouin; Wedi, Nils, and Baousis, Vasileios. Machine learning at ecmwf: A roadmap for the next 10 years, 01/2021 2021. URL https://www.ecmwf.int/node/19877. (Cited on page 43.)

Easterbrook, Steve M. *Computing the Climate: How We Know What We Know About Climate Change*. Cambridge University Press, 2023. (Cited on page 2.)

Eisfelder, Christina; Asam, Sarah; Hirner, Andreas; Reiners, Philipp; Holzwarth, Stefanie; Bachmann, Martin; Gessner, Ursula; Dietz, Andreas; Huth, Juliane; Bachofer, Felix, and Kuenzer, Claudia. Seasonal vegetation trends for europe over 30 years from a novel normalised difference vegetation index (ndvi) time-series-the timeline ndvi product. *Remote Sensing*, 15(14), 2023. ISSN 2072-4292. doi: 10.3390/rs15143616. URL https://www.mdpi.com/2072-4292/15/14/3616. (Cited on page 93.)

Ek, M. B. and Holtslag, A. A. M. Influence of soil moisture on boundary layer cloud development. *Journal of Hydrometeorology*, 5(1):86 – 99, 2004. doi: 10.1175/1525-7541(2004)005<0086:IOSMOB>2.0.CO;2. URL https://journals.ametsoc.org/view/journals/hydr/5/1/1525-7541_2004_005_0086_iosmob_2_0_co_2.xml. (Cited on page 108.)

Elman, Jeffrey L. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990. doi: https://doi.org/10.1207/s15516709cog1402_1. URL https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog1402_1. (Cited on page 14.)

Emerton, Rebecca E.; Stephens, Elisabeth M.; Pappenberger, Florian; Pagano, Thomas C.; Weerts, Albrecht H.; Wood, Andy W.; Salamon, Peter; Brown, James D.; Hjerdt, Niclas; Donnelly, Chantal; Baugh, Calum A., and Cloke, Hannah L. Continental and global scale flood forecasting systems. *WIREs Water*, 3(3):391–418, 2016. doi: https://doi.org/10.1002/wat2.1137. URL https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wat2.1137. (Cited on page 155.)

Entekhabi, Dara; Njoku, Eni G.; O'Neill, Peggy E.; Kellogg, Kent H.; Crow, Wade T.; Edelstein, Wendy N.; Entin, Jared K.; Goodman, Shawn D.; Jackson, Thomas J.; Johnson, Joel; Kimball, John; Piepmeier, Jeffrey R.; Koster, Randal D.; Martin, Neil; McDonald, Kyle C.; Moghaddam, Mahta; Moran, Susan; Reichle, Rolf; Shi, J. C.; Spencer, Michael W.; Thurman, Samuel W.; Tsang, Leung, and Van Zyl, Jakob. The soil moisture active passive (smap) mission. *Proceedings of the IEEE*, 98(5):704–716, 2010. doi: 10.1109/JPROC.2010.2043918. (Cited on page 141.)

Essa, Yassmin H; Hirschi, Martin; Thiery, Wim; El-Kenawy, Ahmed M, and Yang, Chunxue. Drought characteristics in mediterranean under future climate change. *npj Climate and Atmospheric Science*, 6(1):133, 2023. doi: 10.1038/s41612-023-00458-4. (Cited on page 79.)

Fare Garnot, Vivien Sainte and Landrieu, Loic. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4852–4861, 2021. doi: 10.1109/ICCV48922.2021.00483. (Cited on page 50.)

Fazry, Lhuqita; Ramadhan, Mgs M. Luthfi, and Jatmiko, Wisnu. Change detection of high-resolution remote sensing images through adaptive focal modulation on hierarchical feature maps. *IEEE Access*, 11:69072–69090, 2023. doi: 10.1109/ACCESS.2023.3292531. (Cited on page 18.)

Feng, Jia-Chang; Hong, Fa-Ting, and Zheng, Wei-Shi. Mist: Multiple instance self-training framework for video anomaly detection. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (CVPR)*, pages 14009–14018, 2021. (Cited on page 45.)

Ferchichi, Aya; Abbes, Ali Ben; Barra, Vincent, and Farah, Imed Riadh. Forecasting vegetation indices from spatio-temporal remotely sensed data using deep learning-based approaches: A systematic literature review. *Ecological Informatics*, 68:101552, 2022. ISSN 1574-9541. doi: 10.1016/j.ecoinf.2022.101552. URL https://www.sciencedirect.com/science/article/pii/S1574954122000012. (Cited on page 46.)

Flach, M.; Gans, F.; Brenning, A.; Denzler, J.; Reichstein, M.; Rodner, E.; Bathiany, S.; Bodesheim, P.; Guanche, Y.; Sippel, S., and Mahecha, M. D. Multivariate anomaly detection for earth observations: a comparison of algorithms and feature extraction techniques. *Earth System Dynamics*, 8 (3):677–696, 2017. doi: 10.5194/esd-8-677-2017. URL https://esd.copernicus.org/articles/8/677/2017/. (Cited on pages 43 and 113.)

Fornacca, Davide; Ren, Guopeng, and Xiao, Wen. Performance of three modis fire products (mcd45a1, mcd64a1, mcd14ml), and esa fire_cci in a mountainous area of northwest yunnan, china, characterized by frequent small fires. *Remote Sensing*, 9(11):1131, 2017. (Cited on page 52.)

Frame, J. M.; Kratzert, F.; Klotz, D.; Gauch, M.; Shalev, G.; Gilon, O.; Qualls, L. M.; Gupta, H. V., and Nearing, G. S. Deep learning rainfall–runoff predictions of extreme events. *Hydrology and Earth System Sciences*, 26(13):3377–3392, 2022. doi: 10.5194/hess-26-3377-2022. URL https://hess.copernicus.org/articles/26/3377/2022/. (Cited on page 155.)

Fukushima, Kunihiko. Visual feature extraction by a multilayered network of analog threshold elements. *IEEE Transactions on Systems Science and Cybernetics*, 5(4):322–333, 1969. doi: 10.1109/TSSC.1969.300225. (Cited on page 13.)

Fukushima, Kunihiko. Neural network model for a mechanism of pattern recognition unaffected by shift in position- neocognitron. *Electron. & Commun. Japan*, 62(10):11–18, 1979. (Cited on page 13.)

Furusho-Percot, C.; Goergen, K.; Hartick, C.; Poshyvailo-Strube, L., and Kollet, S. Groundwater model impacts multiannual simulations of heat waves. *Geophysical Research Letters*, 49(10):e2021GL096781, 2022. doi: 10.1029/2021GL096781. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GL096781. e2021GL096781 2021GL096781. (Cited on page 79.)

Furusho-Percot, Carina; Goergen, Klaus; Hartick, Carl; Kulkarni, Ketan; Keune, Jessica, and Kollet, Stefan. Pan-european groundwater to atmosphere terrestrial systems climatology from a physically consistent simulation. *Scientific Data*, 6(1):320, 2019. ISSN 2052-4463. doi: 10.1038/s41597-019-0328-7. URL https://doi.org/10.1038/s41597-019-0328-7. (Cited on pages 21, 22, 79 and 93.)

Furusho-Percot, Carina; Goergen, Klaus; Keune, Jessica; Kulkarni, Ketan, and Kollet, Stefan. Pan-european, physically consistent simulations from groundwater to the atmosphere with the Terrestrial Systems Modeling Platform, TerrSysMP (1989-2018 daily time-series), 2019a. URL https://datapub.fz-juelich.de. (Cited on page 223.)

Furusho-Percot, Carina; Goergen, Klaus; Keune, Jessica; Kulkarni, Ketan, and Kollet, Stefan. Pan-european, physically consistent simulations from groundwater to the atmosphere with the Terrestrial Systems Modeling Platform, TerrSysMP (1989-2018 daily time-series), 2019b. URL https://doi.org/10.1594/PANGAEA.901823. (Cited on page 223.)

Gaetano, Raffaele; Ienco, Dino; Ose, Kenji, and Cresson, Remi. A two-branch cnn architecture for land cover classification of pan and ms imagery. *Remote Sensing*, 10(11), 2018. ISSN 2072-4292. doi: 10.3390/rs10111746. URL https://www.mdpi.com/2072-4292/10/11/1746. (Cited on page 49.)

Gan, Chenquan; Yan, Xiaoqin; Wu, Yunfeng, and Zhang, Zufan. A two-branch convolution residual network for image compressive sensing. *IEEE Access*, 8:1705–1714, 2020. doi: 10.1109/ACCESS.2019.2961369. (Cited on page 49.)

Gao, Liang; Liu, Hui; Yang, Minhang; Chen, Long; Wan, Yaling; Xiao, Zhengqing, and Qian, Yurong. Stransfuse: Fusing swin transformer and convolutional neural network for remote sensing image semantic segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:10990–11003, 2021. doi: 10.1109/JSTARS.2021.3119654. (Cited on page 88.)

Gao, Peiqiang; Du, Wenfeng; Lei, Qingwen; Li, Juezhi; Zhang, Shuai, and Li, Ning. Ndvi forecasting model based on the combination of time series decomposition and cnn – lstm. *Water Resources Management*, 37:1481–1497, 2023a. doi: 10.1007/s11269-022-03419-3. (Cited on page 104.)

Gao, Zhihan; Shi, Xingjian; Han, Boran; Wang, Hao; Jin, Xiaoyong; Maddix, Danielle; Zhu, Yi; Li, Mu, and Wang, Yuyang Bernie. Prediff: Precipitation nowcasting with latent diffusion models. *Advances in Neural Information Processing Systems*, 36:78621–78656, 2023b. (Cited on page 2.)

García-García, Almudena; Cuesta-Valero, Francisco José; Miralles, Diego G; Mahecha, Miguel D; Quaas, Johannes; Reichstein, Markus; Zscheischler, Jakob, and Peng, Jian. Soil heat extremes can outpace air temperature extremes. *Nature Climate Change*, 13(11):1237–1241, 2023. (Cited on page 140.)

Garnot, Vivien Sainte Fare and Landrieu, Loic. Lightweight temporal self-attention for classifying satellite images time series. In *Advanced Analytics and Learning on Temporal Data*, pages 171–181, Cham, 2020. Springer International Publishing. ISBN 978-3-030-65742-0. (Cited on page 50.)

Garnot, Vivien Sainte Fare; Landrieu, Loic; Giordano, Sebastien, and Chehata, Nesrine. Satellite image time series classification with pixel-set encoders and temporal self-attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12325–12334, 2020. (Cited on page 50.)

Gasper, F.; Goergen, K.; Shrestha, P.; Sulis, M.; Rihani, J.; Geimer, M., and Kollet, S. Implementation and scaling of the fully coupled terrestrial systems modeling platform (terrsysmp v1.0) in a massively parallel supercomputing environment – a case study on juqueen (ibm blue gene/q). *Geoscientific Model Development*, 7(5):2531–2543, 2014. doi: 10.5194/gmd-7-2531-2014. URL https://gmd.copernicus.org/articles/7/2531/2014/. (Cited on pages 21 and 79.)

Gauch, Martin; Kratzert, Frederik; Gilon, Oren; Gupta, Hoshin; Mai, Juliane; Nearing, Grey; Tolson, Bryan; Hochreiter, Sepp, and Klotz, Daniel. In defense of metrics: Metrics sufficiently encode typical human preferences regarding hydrological model performance. *Water Resources Research*, 59(6):e2022WR033918, 2023. doi: https://doi.org/10.1029/2022WR033918. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022WR033918. e2022WR033918 2022WR033918. (Cited on page 171.)

Geirinhas, JL; Russo, AC; Libonati, R; Miralles, DG; Ramos, AM; Gimeno, L, and Trigo, RM. Combined large-scale tropical and subtropical forcing on the severe 2019–2022 drought in south america. *npj Climate and Atmospheric Science*, 6(1):185, 2023. (Cited on page 108.)

Geiss, S.; Scheck, L.; de Lozar, A., and Weissmann, M. Understanding the model representation of clouds based on visible and infrared satellite observations. *Atmospheric Chemistry and Physics*, 21(16):12273–12290, 2021. doi: 10.5194/acp-21-12273-2021. URL https://acp.copernicus.org/articles/21/12273/2021/. (Cited on pages 80 and 104.)

Gerard, Sebastian; Zhao, Yu, and Sullivan, Josephine. Wildfirespreadts: A dataset of multimodal time series for wildfire spread prediction. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M., and Levine, S., editors, *Advances in Neural Information Processing Systems*, volume 36, pages 74515–74529. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/

ebd545176bdaa9cd5d45954947bd74b7-Paper-Datasets_and_Benchmarks.pdf. (Cited on page 213.)

Gholami, Shahrzad; Kodandapani, Narendran; Wang, Jane, and Ferres, Juan M. Lavista. Where there's smoke, there's fire: Wildfire risk predictive modeling via historical climate data. In *AAAI Conference on Artificial Intelligence*, 2020. (Cited on page 45.)

Giannaros, Theodore M.; Papavasileiou, Georgios; Lagouvardos, Konstantinos; Kotroni, Vassiliki; Dafis, Stavros; Karagiannidis, Athanasios, and Dragozi, Eleni. Meteorological analysis of the 2021 extreme wildfires in greece: Lessons learned and implications for early warning of the potential for pyroconvection. *Atmosphere*, 13(3), 2022. ISSN 2073-4433. doi: 10.3390/atmos13030475. URL https://www.mdpi.com/2073-4433/13/3/475. (Cited on page 20.)

Giglio, Louis; Schroeder, Wilfrid, and Justice, Christopher O. The collection 6 modis active fire detection algorithm and fire products. *Remote Sensing of Environment*, 178:31–41, 2016. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2016.02.054. URL https://www.sciencedirect.com/science/article/pii/S0034425716300827. (Cited on page 20.)

Giorgi, F.; Jones, C., and Asrar, G. R. Addressing climate information needs at the regional level: the cordex framework. *Bulletin - World Meteorological Organization*, 58(3):175–183, 2009. (Cited on pages 21, 27, 28 and 116.)

Giuseppe, Francesca Di; Pappenberger, Florian; Wetterhall, Fredrik; Krzeminski, Blazej; Camia, Andrea; Libertá, Giorgio, and Miguel, Jesus San. The potential predictability of fire danger provided by numerical weather prediction. *Journal of Applied Meteorology and Climatology*, 55(11):2469 – 2491, 2016. doi: 10.1175/JAMC-D-15-0297.1. URL https://journals.ametsoc.org/view/journals/apme/55/11/jamc-d-15-0297.1.xml. (Cited on page 45.)

Graves, Alex. Generating sequences with recurrent neural networks. *arXiv preprint arXiv:1308.0850*, 2013. (Cited on page 14.)

Green, J.; Haigh, I. D.; Quinn, N.; Neal, J.; Wahl, T.; Wood, M.; Eilander, D.; de Ruiter, M.; Ward, P., and Camus, P. Review article: A comprehensive review of compound flooding literature with a focus on coastal and estuarine regions. *Natural Hazards and Earth System Sciences*, 25(2):747–816, 2025. doi: 10.5194/nhess-25-747-2025. URL https://nhess.copernicus.org/articles/25/747/2025/. (Cited on page 47.)

Gu, Albert and Dao, Tri. Mamba: Linear-time sequence modeling with selective state spaces. In *First Conference on Language Modeling*, 2024. (Cited on pages 15, 18, 19, 111, 128, 130, 131, 134, 155 and 158.)

Gu, Albert; Johnson, Isys; Goel, Karan; Saab, Khaled; Dao, Tri; Rudra, Atri, and Ré, Christopher. Combining recurrent, convolutional, and continuous-time models with linear state space layers. In *Advances in Neural Information Processing Systems*, volume 34, pages 572–585. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/05546b0e38ab9175cd905eebcc6ebb76-Paper.pdf. (Cited on pages 18 and 155.)

Gu, Albert; Goel, Karan, and Ré, Christopher. Efficiently modeling long sequences with structured state spaces. In *The International Conference on Learning Representations (ICLR)*, 2022. (Cited on pages 18, 19 and 155.)

Guo, Zifeng; Leitão, João P.; Simões, Nuno E., and Moosavi, Vahid. Data-driven flood emulation: Speeding up urban flood predictions by deep convolutional neural networks. *Journal of Flood Risk Management*, 14(1):e12684, 2021. doi: https://doi.org/10.1111/jfr3.12684. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jfr3.12684. (Cited on page 47.)

Gutowski Jr., W. J.; Giorgi, F.; Timbal, B.; Frigon, A.; Jacob, D.; Kang, H.-S.; Raghavan, K.; Lee, B.; Lennard, C.; Nikulin, G.; O'Rourke, E.; Rixen, M.; Solman, S.; Stephenson, T., and Tangang, F. Wcrp coordinated regional downscaling experiment (cordex): a diagnostic mip for cmip6. *Geoscientific Model Development*, 9(11):4087–4095, 2016. doi: 10.5194/gmd-9-4087-2016. URL https://gmd.copernicus.org/articles/9/4087/2016/. (Cited on page 21.)

Hammad, Ahmed T. and Falchetta, Giacomo. Probabilistic forecasting of remotely sensed cropland vegetation health and its relevance for food security. *Science of The Total Environment*, 838:156157, 2022. ISSN 0048-9697. doi: 10.1016/j.scitotenv.2022.156157. URL https://www.sciencedirect.com/science/article/pii/S0048969722032545. (Cited on page 46.)

Han, Songqiao; Hu, Xiyang; Huang, Hailiang; Jiang, Minqi, and Zhao, Yue. ADBench: Anomaly detection benchmark. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022. URL https://openreview.net/forum?id=foA_SFQ9zo0. (Cited on page 44.)

Hantson, Stijn; Arneth, Almut; Harrison, Sandy P; Kelley, Douglas I; Prentice, I Colin; Rabin, Sam S; Archibald, Sally; Mouillot, Florent; Arnold, Steve R; Artaxo, Paulo, and others, . The status and challenge of global fire modelling. *Biogeosciences*, 13(11):3359–3375, 2016. (Cited on pages 52 and 58.)

Hao, Zengchao; Singh, Vijay P., and Xia, Youlong. Seasonal drought prediction: Advances, challenges, and future prospects. *Reviews of Geophysics*, 56(1):108–141, 2018. doi: https://doi.org/10.1002/2016RG000549. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016RG000549. (Cited on pages 24 and 108.)

Hapuarachchi, H. A. P.; Bari, M. A.; Kabir, A.; Hasan, M. M.; Woldemeskel, F. M.; Gamage, N.; Sunter, P. D.; Zhang, X. S.; Robertson, D. E.; Bennett, J. C., and Feikema, P. M. Development of a national 7-day ensemble streamflow forecasting service for australia. *Hydrology and Earth System Sciences*, 26(18):4801–4821, 2022. doi: 10.5194/hess-26-4801-2022. URL https://hess.copernicus.org/articles/26/4801/2022/. (Cited on page 155.)

Harrigan, S.; Zsoter, E.; Alfieri, L.; Prudhomme, C.; Salamon, P.; Wetterhall, F.; Barnard, C.; Cloke, H., and Pappenberger, F. Glofas-era5 operational global river discharge reanalysis 1979–present. *Earth System Science Data*, 12(3):2043–2060, 2020. doi: 10.5194/essd-12-2043-2020. URL https://essd.copernicus.org/articles/12/2043/2020/. (Cited on pages 30, 31, 32, 166, 181 and 224.)

Harrigan, S.; Zsoter, E.; Cloke, H.; Salamon, P., and Prudhomme, C. Daily ensemble river discharge reforecasts and real-time forecasts from the operational global flood awareness system. *Hydrology and Earth System Sciences*, 27(1):1–19, 2023. doi: 10.5194/hess-27-1-2023. URL https://hess.copernicus.org/articles/27/1/2023/. (Cited on pages 32, 155, 167, 169, 195, 208 and 218.)

Hauer, C; Paster, M; Pulg, U; Ofenböck, T, and Habersack, H. Critical flows at the wien river during the 1000-years event in september 2024–causes, consequences and possible management options for urban river flood management. *Natural Hazards*, pages 1–13, 2025. doi: https://doi.org/10.1007/s11069-025-07244-w. (Cited on page 200.)

He, Haoyang; Bai, Yuhu; Zhang, Jiangning; He, Qingdong; Chen, Hongxu; Gan, Zhenye; Wang, Chengjie; Li, Xiangtai; Tian, Guanzhong, and Xie, Lei. Mambaad: Exploring state space models for multi-class unsupervised anomaly detection. *Advances in Neural Information Processing Systems*, 37:71162–71187, 2024. (Cited on page 44.)

Hendrycks, Dan and Gimpel, Kevin. Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415*, 2016. (Cited on pages 84 and 207.)

Herring, Stephanie C.; Christidis, Nikolaos; Hoell, Andrew, and Stott, Peter A. Explaining extreme events of 2020 from a climate perspective. *Bulletin of the American Meteorological Society*, 103(3):S1 – S129, 2022. doi: 10.1175/BAMS-ExplainingExtremeEvents2020.1. URL https://journals.ametsoc.org/view/journals/bams/103/3/BAMS-ExplainingExtremeEvents2020.1.xml. (Cited on page 108.)

Hersbach, Hans; Bell, Bill; Berrisford, Paul; Hirahara, Shoji; Horányi, András; Muñoz-Sabater, Joaquín; Nicolas, Julien; Peubey, Carole; Radu, Raluca; Schepers, Dinand; Simmons, Adrian; Soci, Cornel; Abdalla, Saleh; Abellan, Xavier; Balsamo, Gianpaolo; Bechtold, Peter; Biavati, Gionata; Bidlot, Jean; Bonavita, Massimo; De Chiara, Giovanna; Dahlgren, Per; Dee, Dick; Diamantakis, Michail; Dragani, Rossana; Flemming, Johannes; Forbes, Richard; Fuentes, Manuel; Geer, Alan; Haimberger, Leo; Healy, Sean; Hogan, Robin J.; Hólm, Elías; Janisková, Marta; Keeley, Sarah; Laloyaux, Patrick; Lopez, Philippe; Lupu, Cristina; Radnoti, Gabor; de Rosnay, Patricia; Rozum, Iryna; Vamborg, Freja; Villaume, Sebastien, and Thépaut, Jean-Noël. The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730):1999–2049, 2020. doi: 10.1002/qj.3803. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803. (Cited on pages 3, 27, 30, 166 and 218.)

Heudorfer, Benedikt; Gupta, Hoshin V., and Loritz, Ralf. Are deep learning models in hydrology entity aware? *Geophysical Research Letters*, 52(6):e2024GL113036, 2025. doi: https://doi.org/10.1029/2024GL113036. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024GL113036. e2024GL113036 2024GL113036. (Cited on page 47.)

Hilbert, David and Hilbert, David. Über die stetige abbildung einer linie auf ein flächenstück. *Dritter Band: Analysis· Grundlagen der Mathematik· Physik Verschiedenes: Nebst Einer Lebensgeschichte*, pages 1–2, 1935. doi: https://doi.org/10.1007/BF01199431. (Cited on page 159.)

Hochman, A.; Marra, F.; Messori, G.; Pinto, J. G.; Raveh-Rubin, S.; Yosef, Y., and Zittis, G. Extreme weather and societal impacts in the eastern mediterranean. *Earth System Dynamics*, 13(2):

749–777, 2022. doi: 10.5194/esd-13-749-2022. URL https://esd.copernicus.org/articles/13/749/2022/. (Cited on page 52.)

Hochreiter, Sepp. Untersuchungen zu dynamischen neuronalen netzen. *Diploma, Technische Universität München*, 91(1):31, 1991. (Cited on page 14.)

Hochreiter, Sepp and Schmidhuber, Jürgen. Long short-term memory. *Neural computation*, 9:1735–80, 12 1997. doi: 10.1162/neco.1997.9.8.1735. (Cited on pages 14, 15, 47, 49 and 59.)

Hofmann, Julian and Schüttrumpf, Holger. floodgan: Using deep adversarial learning to predict pluvial flooding in real time. *Water*, 13(16), 2021. ISSN 2073-4441. doi: 10.3390/w13162255. URL https://www.mdpi.com/2073-4441/13/16/2255. (Cited on page 47.)

Horn, B.K.P. Hill shading and the reflectance map. *Proceedings of the IEEE*, 69(1):14–47, 1981. doi: 10.1109/PROC.1981.11918. (Cited on page 22.)

Hu, Jie; Shen, Li, and Sun, Gang. Squeeze-and-excitation networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7132–7141, 2018. doi: 10.1109/CVPR.2018.00745. (Cited on page 86.)

Hu, Vincent Tao; Baumann, Stefan Andreas; Gui, Ming; Grebenkova, Olga; Ma, Pingchuan; Fischer, Johannes, and Ommer, Björn. Zigma: A dit-style zigzag mamba diffusion model. In *European Conference on Computer Vision (ECCV) 2024*, pages 148–166, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-72664-4. doi: https://doi.org/10.1007/978-3-031-72664-4_9. (Cited on page 50.)

Huang, Chaoqin; Xu, Qinwei; Wang, Yanfeng; Wang, Yu, and Zhang, Ya. Self-supervised masking for unsupervised anomaly detection and localization. *IEEE Transactions on Multimedia*, 25:4426–4438, 2023. doi: 10.1109/TMM.2022.3175611. (Cited on page 44.)

Huang, Shengzhi; Huang, Qiang; Chang, Jianxia; Leng, Guoyong, and Xing, Li. The response of agricultural drought to meteorological drought and the influencing factors: A case study in the wei river basin, china. *Agricultural Water Management*, 159:45–54, 2015. ISSN 0378-3774. doi: https://doi.org/10.1016/j.agwat.2015.05.023. URL https://www.sciencedirect.com/science/article/pii/S0378377415300147. (Cited on pages 24, 43 and 108.)

Huang, Tao; Pei, Xiaohuan; You, Shan; Wang, Fei; Qian, Chen, and Xu, Chang. Localmamba: Visual state space model with windowed selective scan. In *European Conference on Computer Vision (ECCV) 2024 Workshops*, pages 12–22, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-91979-4. doi: 10.1007/978-3-031-91979-4_2. (Cited on page 50.)

Huang, Xun and Belongie, Serge J. Arbitrary style transfer in real-time with adaptive instance normalization. *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 1510–1519, 2017. (Cited on page 48.)

Huot, Fantine; Hu, R. Lily; Goyal, Nita; Sankar, Tharun; Ihme, Matthias, and Chen, Yi-Fan. Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2022. doi: 10.1109/TGRS.2022.3192974. (Cited on pages 19 and 45.)

Iban, Muzaffer Can and Sekertekin, Aliihsan. Machine learning based wildfire susceptibility mapping using remotely sensed fire data and gis: A case study of adana and mersin provinces, turkey. *Ecological Informatics*, 69:101647, 2022. ISSN 1574-9541. doi: https://doi.org/10.1016/j.ecoinf.2022.101647. URL https://www.sciencedirect.com/science/article/pii/S1574954122000966. (Cited on page 45.)

Ioffe, Sergey and Szegedy, Christian. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015. (Cited on pages 48 and 55.)

Jacob, Daniela; Teichmann, Claas; Sobolowski, Stefan; Katragkou, Eleni; Anders, Ivonne; Belda, Michal; Benestad, Rasmus; Boberg, Fredrik; Buonomo, Erasmo; Cardoso, Rita M; Casanueva, Ana; Christensen, Ole B.; Christensen, Jens Hesselbjerg; Coppola, Erika; De Cruz, Lesley; Davin, Edouard L.; Dobler, Andreas; Domínguez, Marta; Fealy, Rowan; Fernandez, Jesus; Gaertner, Miguel Angel; García-Díez, Markel; Giorgi, Filippo; Gobiet, Andreas; Goergen, Klaus; Gómez-Navarro, Juan José; Alemán, Juan Jesús González; Gutiérrez, Claudia; Gutiérrez, José M.; Güttler, Ivan; Haensler, Andreas; Halenka, Tomáš; Jerez, Sonia; Jiménez-Guerrero, Pedro; Jones, Richard G.; Keuler, Klaus; Kjellström, Erik; Knist, Sebastian; Kotlarski, Sven; Maraun, Douglas; van Meijgaard, Erik; Mercogliano, Paola; Montávez, Juan Pedro; Navarra, Antonio; Nikulin, Grigory; de Noblet-Ducoudré, Nathalie; Panitz, Hans-Juergen; Pfeifer, Susanne; Piazza, Marie; Pichelli, Emanuela; Pietikäinen, Joni-Pekka; Prein, Andreas F.; Preuschmann, Swantjen; Rechid, Diana; Rockel, Burkhardt; Romera, Raquel; Sánchez, Enrique; Sieck, Kevin; Soares, Pedro M. M.; Somot, Samuel; Srnec, Lidija; Sørland, Silje Lund; Termonia, Piet; Truhetz, Heimo; Vautard, Robert; Warrach-Sagi, Kirsten, and Wulfmeyer, Volker. Regional climate downscaling over europe: perspectives from the euro-cordex community. *Regional Environmental Change*, 20:51, 04 2020. doi: 10.1007/s10113-020-01606-9. (Cited on page 21.)

Jain, Piyush; Coogan, Sean CP; Subramanian, Sriram Ganapathi; Crowley, Mark; Taylor, Steve, and Flannigan, Mike D. A review of machine learning applications in wildfire science and management. *Environmental Reviews*, 28(4):478–505, 2020. (Cited on pages 45 and 52.)

Jakoel, Karin; Efraim, Liron, and Shaham, Tamar Rott. Gans spatial control via inference-time adaptive normalization. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 31–40, 2022. doi: 10.1109/WACV51458.2022.00011. (Cited on page 48.)

Janizadeh, Saeid; Chandra Pal, Subodh; Saha, Asish; Chowdhuri, Indrajit; Ahmadi, Kourosh; Mirzaei, Sajjad; Mosavi, Amir Hossein, and Tiefenbacher, John P. Mapping the spatial and temporal variability of flood hazard affected by climate and land-use changes in the future. *Journal of Environmental Management*, 298:113551, 2021. ISSN 0301-4797. doi: https://doi.org/10.1016/j.jenvman.2021.113551. URL https://www.sciencedirect.com/science/article/pii/S0301479721016133. (Cited on page 47.)

Jansen, Aren; Ellis, Daniel PW; Hershey, Shawn; Moore, R Channing; Plakal, Manoj; Popat, Ashok C, and Saurous, Rif A. Coincidence, categorization, and consolidation: Learning to recognize sounds with minimal supervision. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 121–125. IEEE, 2020. (Cited on page 112.)

Jefferson, Jennifer L. and Maxwell, Reed M. Evaluation of simple to complex parameterizations of bare ground evaporation. *Journal of Advances in Modeling Earth Systems*, 7(3):1075–1092, 2015. doi: 10.1002/2014MS000398. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014MS000398. (Cited on page 21.)

Ji, Yuheng; Wang, Dan; Li, Qingliang; Liu, Taihui, and Bai, Yu. Global wildfire danger predictions based on deep learning taking into account static and dynamic variables. *Forests*, 15(1), 2024. ISSN 1999-4907. doi: 10.3390/f15010216. URL https://www.mdpi.com/1999-4907/15/1/216. (Cited on page 63.)

Jia, Donghuai; Li, Weide; Huang, Di, and Chen, Shukai. Daily runoff prediction based on lightweight mamba with partial normalization. *Hydrology Research*, 55(12):1182–1196, 11 2024. ISSN 0029-1277. doi: 10.2166/nh.2024.063. URL https://doi.org/10.2166/nh.2024.063. (Cited on page 50.)

Jiang, Minqi; Hou, Chaochuan; Zheng, Ao; Hu, Xiyang; Han, Songqiao; Huang, Hailiang; He, Xiangnan; Yu, Philip S, and Zhao, Yue. Weakly supervised anomaly detection: A survey. *arXiv preprint arXiv:2302.04549*, 2023. (Cited on page 45.)

Jiang, S.; Bevacqua, E., and Zscheischler, J. River flooding mechanisms and their changes in europe revealed by explainable machine learning. *Hydrology and Earth System Sciences*, 26(24):6339–6359, 2022a. doi: 10.5194/hess-26-6339-2022. URL https://hess.copernicus.org/articles/26/6339/2022/. (Cited on page 155.)

Jiang, Shijie; Tarasova, Larisa; Yu, Guo, and Zscheischler, Jakob. Compounding effects in flood drivers challenge estimates of extreme river floods. *Science Advances*, 10(13):eadl4005, 2024. doi: 10.1126/sciadv.adl4005. URL https://www.science.org/doi/abs/10.1126/sciadv.adl4005. (Cited on page 155.)

Jiang, Tanqiu; Bendre, Sidhant K.; Lyu, Hanjia, and Luo, Jiebo. From static to dynamic prediction: Wildfire risk assessment based on multiple environmental factors. In *IEEE International Conference on Big Data*, pages 4877–4886, 2021. doi: 10.1109/BigData52589.2021.9672044. (Cited on page 45.)

Jiang, Tao; Xie, Weiying; Li, Yunsong; Lei, Jie, and Du, Qian. Weakly supervised discriminative learning with spectral constrained generative adversarial network for hyperspectral anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems*, 33(11):6504–6517, 2022b. doi: 10.1109/TNNLS.2021.3082158. (Cited on page 45.)

Jin, Hongxiao; Vicente-Serrano, Sergio M; Tian, Feng; Cai, Zhanzhang; Conradt, Tobias; Boincean, Boris; Murphy, Conor; Farizo, Begoña Alvarez; Grainger, Sam; López-Moreno, Juan I, and Eklundh, Lars. Higher vegetation sensitivity to meteorological drought in autumn than spring across european biomes. *Communications Earth & Environment*, 4(1):299, 2023. doi: 10.1038/s43247-023-00960-w. (Cited on page 79.)

Johnson, Justin; Alahi, Alexandre, and Fei-Fei, Li. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision (ECCV) 2016*, pages 694–711,

Cham, 2016. Springer International Publishing. doi: 10.1007/978-3-319-46475-6_43. (Cited on page 87.)

Jones, Anne; Kuehnert, Julian; Fraccaro, Paolo; Meuriot, Ophélie; Ishikawa, Tatsuya; Edwards, Blair; Stoyanov, Nikola; Remy, Sekou L; Weldemariam, Kommy, and Assefa, Solomon. Ai for climate impacts: applications in flood risk. *npj Climate and Atmospheric Science*, 6(1):63, 2023. doi: https://doi.org/10.1038/s41612-023-00388-1. (Cited on page 155.)

Jones, Jim E. and Woodward, Carol S. Newton–krylov-multigrid solvers for large-scale, highly heterogeneous, variably saturated flow problems. *Advances in Water Resources*, 24(7):763–774, 2001. ISSN 0309-1708. doi: 10.1016/S0309-1708(00)00075-0. URL https://www.sciencedirect.com/science/article/pii/S0309170800000750. (Cited on page 21.)

Jones, Philip W. First- and second-order conservative remapping schemes for grids in spherical coordinates. *Monthly Weather Review*, 127(9):2204 – 2210, 1999. doi: 10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2. URL https://journals.ametsoc.org/view/journals/mwre/127/9/1520-0493_1999_127_2204_fasocr_2.0.co_2.xml. (Cited on page 26.)

Kalnay, Eugenia; Mote, Safa, and Da, Cheng. *Earth System Modeling, Data Assimilation and Predictability: Atmosphere, Oceans, Land and Human Systems*. Cambridge University Press, 2 edition, 2024. (Cited on page 2.)

Karasante, Ilektra; Alonso, Lazaro; Prapas, Ioannis; Ahuja, Akanksha; Carvalhais, Nuno, and Papoutsis, Ioannis. Seasfire cube-a multivariate dataset for global wildfire modeling. *Scientific Data*, 12(1):368, 2025. doi: 10.1038/s41597-025-04546-3. (Cited on page 213.)

Karim, Fazlul; Armin, Mohammed Ali; Ahmedt-Aristizabal, David; Tychsen-Smith, Lachlan, and Petersson, Lars. A review of hydrodynamic and machine learning approaches for flood inundation modeling. *Water*, 15(3), 2023. ISSN 2073-4441. doi: 10.3390/w15030566. URL https://www.mdpi.com/2073-4441/15/3/566. (Cited on page 47.)

Kazadi, Arnold; Doss-Gollin, James; Sebastian, Antonia, and Silva, Arlei. Floodgnn-gru: a spatio-temporal graph neural network for flood prediction. *Environmental Data Science*, 3:e21, 2024. doi: 10.1017/eds.2024.19. (Cited on page 47.)

Kew, S. F.; Philip, S. Y.; Hauser, M.; Hobbins, M.; Wanders, N.; van Oldenborgh, G. J.; van der Wiel, K.; Veldkamp, T. I. E.; Kimutai, J.; Funk, C., and Otto, F. E. L. Impact of precipitation and increasing temperatures on drought trends in eastern africa. *Earth System Dynamics*, 12(1):17–35, 2021. doi: 10.5194/esd-12-17-2021. URL https://esd.copernicus.org/articles/12/17/2021/. (Cited on page 79.)

Kingma, Diederik P and Ba, Jimmy. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. (Cited on pages 58 and 116.)

Kirschstein, Nikolas and Sun, Yixuan. The merit of river network topology for neural flood forecasting. In *International Conference on Machine Learning*, pages 24713–24725. PMLR, 2024. (Cited on page 47.)

Kladny, Klaus-Rudolf; Milanta, Marco; Mraz, Oto; Hufkens, Koen, and Stocker, Benjamin D. Deep learning for satellite image forecasting of vegetation greenness. *bioRxiv*, 2022. doi: 10.1101/2022.08.16.504173. URL https://www.biorxiv.org/content/early/2022/08/17/2022.08.16.504173. (Cited on page 46.)

Klemmer, Konstantin; Rolf, Esther; Robinson, Caleb; Mackey, Lester, and Rußwurm, Marc. Satclip: Global, general-purpose location embeddings with satellite imagery. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(4):4347–4355, Apr. 2025. doi: 10.1609/aaai.v39i4.32457. URL https://ojs.aaai.org/index.php/AAAI/article/view/32457. (Cited on page 220.)

Klimavičius, Laurynas; Rimkus, Egidijus; Stonevičius, Edvinas, and Mačiulytė, Viktorija. Seasonality and long-term trends of ndvi values in different land use types in the eastern part of the baltic sea basin. *Oceanologia*, 65(1):171–181, 2023. ISSN 0078-3234. doi: 10.1016/j.oceano.2022.02.007. URL https://www.sciencedirect.com/science/article/pii/S0078323422000240. (Cited on page 101.)

Knijff, J. M. Van Der; Younis, J., and and, A. P. J. De Roo. Lisflood: a gis-based distributed model for river basin scale water balance and flood simulation. *International Journal of Geographical Information Science*, 24(2):189–212, 2010. doi: 10.1080/13658810802549154. URL https://doi.org/10.1080/13658810802549154. (Cited on pages 30 and 166.)

Kogan, F. N. Remote sensing of weather impacts on vegetation in non-homogeneous areas. *International Journal of Remote Sensing*, 11(8):1405–1419, 1990. doi: 10.1080/01431169008955102. URL https://doi.org/10.1080/01431169008955102. (Cited on page 25.)

Kogan, Felix; Guo, Wei, and Jelenak, Aleksandar. Global vegetation health: Long-term data records. In *Use of Satellite and In-Situ Data to Improve Sustainability*, pages 247–255, Dordrecht, 2011. Springer Netherlands. ISBN 978-90-481-9618-0. doi: 10.1007/978-90-481-9618-0_28. (Cited on page 23.)

Kogan, Felix; Goldberg, Mitch; Schott, Tom, and Guo, Wei. Suomi npp/viirs: improving drought watch, crop loss prediction, and food security. *International Journal of Remote Sensing*, 36(21):5373–5383, 2015. doi: 10.1080/01431161.2015.1095370. URL https://doi.org/10.1080/01431161.2015.1095370. (Cited on pages 25 and 98.)

Kogan, Felix; Guo, Wei, and Yang, Wenze. Near 40-year drought trend during 1981-2019 earth warming and food security. *Geomatics, Natural Hazards and Risk*, 11(1):469–490, 2020. doi: 10.1080/19475705.2020.1730452. URL https://doi.org/10.1080/19475705.2020.1730452. (Cited on pages 79 and 100.)

Kogan, Felix N. Droughts of the late 1980s in the united states as derived from noaa polar-orbiting satellite data. *Bulletin of the American Meteorological Society*, 76(5):655 – 668, 1995a. doi: 10.1175/1520-0477(1995)076<0655:DOTLIT>2.0.CO;2. URL https://journals.ametsoc.org/view/journals/bams/76/5/1520-0477_1995_076_0655_dotlit_2_0_co_2.xml. (Cited on pages 23 and 93.)

Kogan, F.N. Application of vegetation index and brightness temperature for drought detection. *Advances in Space Research*, 15(11):91–100, 1995b. ISSN 0273-1177. doi: 10.1016/0273-1177(95)00079-T. URL https://www.sciencedirect.com/science/article/pii/027311779500079T. Natural Hazards: Monitoring and Assessment Using Remote Sensing Technique. (Cited on pages 25 and 80.)

Kokhlikyan, Narine; Miglani, Vivek; Martin, Miguel; Wang, Edward; Alsallakh, Bilal; Reynolds, Jonathan; Melnikov, Alexander; Kliushkina, Natalia; Araya, Carlos; Yan, Siqi, and others, . Captum: A unified and generic model interpretability library for pytorch. *arXiv preprint arXiv:2009.07896*, 2020. (Cited on page 117.)

Kollet, Stefan J. and Maxwell, Reed M. Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. *Advances in Water Resources*, 29(7):945–958, 2006. ISSN 0309-1708. doi: 10.1016/j.advwatres.2005.08.006. URL https://www.sciencedirect.com/science/article/pii/S0309170805002101. (Cited on page 21.)

Kondylatos, Spyridon; Prapas, Ioannis; Camps-Valls, Gustau, and Papoutsis, Ioannis. Mesogeos: A multi-purpose dataset for data-driven wildfire modeling in the mediterranean. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M., and Levine, S., editors, *Advances in Neural Information Processing Systems*, volume 36, pages 50661–50676. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/9ee3ed2dd656402f954ef9dc37e39f48-Paper-Datasets_and_Benchmarks.pdf. (Cited on page 213.)

Kondylatos, Spyros; Prapas, Ioannis; Ronco, Michele; Papoutsis, Ioannis; Camps-Valls, Gustau; Piles, María; Fernández-Torres, Miguel-Ángel, and Carvalhais, Nuno. Wildfire danger prediction and understanding with deep learning. *Geophysical Research Letters*, 49 (17):e2022GL099368, 2022. doi: https://doi.org/10.1029/2022GL099368. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022GL099368. (Cited on pages 19, 20, 21, 45, 46, 52, 53, 59, 60, 61, 62, 63 and 69.)

Kracheletz, Magdalena; Liu, Ziyu; Springer, Anne; Kusche, Jürgen, and Friederichs, Petra. Would the 2021 western europe flood event be visible in satellite gravimetry? *Journal of Geophysical Research: Atmospheres*, 130(3):e2024JD042190, 2025. doi: https://doi.org/10.1029/2024JD042190. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024JD042190. e2024JD042190 2024JD042190. (Cited on page 198.)

Kraft, Basil; Jung, Martin; Körner, Marco; Requena Mesa, Christian; Cortés, José, and Reichstein, Markus. Identifying dynamic memory effects on vegetation state using recurrent neural networks. *Frontiers in Big Data*, 2, 2019. ISSN 2624-909X. doi: 10.3389/fdata.2019.00031. URL https://www.frontiersin.org/articles/10.3389/fdata.2019.00031. (Cited on page 46.)

Kratzert, F.; Klotz, D.; Shalev, G.; Klambauer, G.; Hochreiter, S., and Nearing, G. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, 23(12):5089–5110, 2019. doi:

10.5194/hess-23-5089-2019. URL https://hess.copernicus.org/articles/23/5089/2019/. (Cited on pages 47, 48 and 215.)

Kratzert, F.; Gauch, M.; Klotz, D., and Nearing, G. Hess opinions: Never train a long short-term memory (lstm) network on a single basin. *Hydrology and Earth System Sciences*, 28(17):4187–4201, 2024. doi: 10.5194/hess-28-4187-2024. URL https://hess.copernicus.org/articles/28/4187/2024/. (Cited on page 47.)

Kreibich, Heidi; Van Loon, Anne F; Schröter, Kai; Ward, Philip J; Mazzoleni, Maurizio; Sairam, Nivedita; Abeshu, Guta Wakbulcho; Agafonova, Svetlana; AghaKouchak, Amir; Aksoy, Hafzullah, and others, . The challenge of unprecedented floods and droughts in risk management. *Nature*, 608(7921):80–86, 2022. doi: https://doi.org/10.1038/s41586-022-04917-5. (Cited on page 155.)

Krizhevsky, Alex; Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012. (Cited on page 14.)

Kuffour, B. N. O.; Engdahl, N. B.; Woodward, C. S.; Condon, L. E.; Kollet, S., and Maxwell, R. M. Simulating coupled surface–subsurface flows with parflow v3.5.0: capabilities, applications, and ongoing development of an open-source, massively parallel, integrated hydrologic model. *Geoscientific Model Development*, 13(3):1373–1397, 2020. doi: 10.5194/gmd-13-1373-2020. URL https://gmd.copernicus.org/articles/13/1373/2020/. (Cited on page 21.)

Lahsen, Myanna and Ribot, Jesse. Politics of attributing extreme events and disasters to climate change. *WIREs Climate Change*, 13(1):e750, 2022. doi: https://doi.org/10.1002/wcc.750. URL https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcc.750. (Cited on page 155.)

Lal, Preet; Singh, Gurjeet; Das, Narendra N.; Colliander, Andreas, and Entekhabi, Dara. Assessment of era5-land volumetric soil water layer product using in situ and smap soil moisture observations. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022. doi: 10.1109/LGRS.2022.3223985. (Cited on page 141.)

Lam, Remi; Sanchez-Gonzalez, Alvaro; Willson, Matthew; Wirnsberger, Peter; Fortunato, Meire; Pritzel, Alexander; Ravuri, Suman; Ewalds, Timo; Alet, Ferran; Eaton-Rosen, Zach; Hu, Weihua; Merose, Alexander; Hoyer, Stephan; Holland, George; Stott, Jacklynn; Vinyals, Oriol; Mohamed, Shakir, and Battaglia, Peter. Graphcast: Learning skillful medium-range global weather forecasting. *arXiv preprint arXiv:2212.12794*, 2022. (Cited on pages 2 and 46.)

Lang, Simon; Alexe, Mihai; Chantry, Matthew; Dramsch, Jesper; Pinault, Florian; Raoult, Baudouin; Clare, Mariana CA; Lessig, Christian; Maier-Gerber, Michael; Magnusson, Linus, and others, . Aifs–ecmwf's data-driven forecasting system. *arXiv preprint arXiv:2406.01465*, 2024a. (Cited on page 219.)

Lang, Simon; Alexe, Mihai; Clare, Mariana CA; Roberts, Christopher; Adewoyin, Rilwan; Boudallègue, Zied Ben; Chantry, Matthew; Dramsch, Jesper; Dueben, Peter D; Hahner, Sara, and others, . Aifs-crps: ensemble forecasting using a model trained with a loss function based on the continuous ranked probability score. *arXiv preprint arXiv:2412.15832*, 2024b. (Cited on page 216.)

Lausier, Anne M and Jain, Shaleen. Overlooked trends in observed global annual precipitation reveal underestimated risks. *Scientific reports*, 8(1):1–7, 2018. (Cited on page 52.)

Lawrence, David M.; Fisher, Rosie A.; Koven, Charles D.; Oleson, Keith W.; Swenson, Sean C.; Bonan, Gordon; Collier, Nathan; Ghimire, Bardan; van Kampenhout, Leo; Kennedy, Daniel; Kluzek, Erik; Lawrence, Peter J.; Li, Fang; Li, Hongyi; Lombardozzi, Danica; Riley, William J.; Sacks, William J.; Shi, Mingjie; Vertenstein, Mariana; Wieder, William R.; Xu, Chonggang; Ali, Ashehad A.; Badger, Andrew M.; Bisht, Gautam; van den Broeke, Michiel; Brunke, Michael A.; Burns, Sean P.; Buzan, Jonathan; Clark, Martyn; Craig, Anthony; Dahlin, Kyla; Drewniak, Beth; Fisher, Joshua B.; Flanner, Mark; Fox, Andrew M.; Gentine, Pierre; Hoffman, Forrest; Keppel-Aleks, Gretchen; Knox, Ryan; Kumar, Sanjiv; Lenaerts, Jan; Leung, L. Ruby; Lipscomb, William H.; Lu, Yaqiong; Pandey, Ashutosh; Pelletier, Jon D.; Perket, Justin; Randerson, James T.; Ricciuto, Daniel M.; Sanderson, Benjamin M.; Slater, Andrew; Subin, Zachary M.; Tang, Jinyun; Thomas, R. Quinn; Val Martin, Maria, and Zeng, Xubin. The community land model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems*, 11(12):4245–4287, 2019. doi: 10.1029/2018MS001583. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001583. (Cited on page 79.)

Le, Hung Van; Hoang, Duc Anh; Tran, Chuyen Trung; Nguyen, Phi Quoc; Tran, Van Hai Thi; Hoang, Nhat Duc; Amiri, Mahdis; Ngo, Thao Phuong Thi; Nhu, Ha Viet; Hoang, Thong Van, and Tien Bui, Dieu. A new approach of deep neural computing for spatial prediction of wildfire danger at tropical climate areas. *Ecological Informatics*, 63:101300, 2021. ISSN 1574-9541. doi: https://doi.org/10.1016/j.ecoinf.2021.101300. URL https://www.sciencedirect.com/science/article/pii/S1574954121000911. (Cited on page 45.)

Lecun, Y.; Bottou, L.; Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791. (Cited on page 13.)

LeCun, Yann; Boser, Bernhard; Denker, John S; Henderson, Donnie; Howard, Richard E; Hubbard, Wayne, and Jackel, Lawrence D. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989. (Cited on page 13.)

LeCun, Yann; Bengio, Yoshua, and Hinton, Geoffrey. Deep learning. *Nature*, 521(7553):436–444, 2015. doi: 10.1038/nature14539. (Cited on page 12.)

Ledig, Christian; Theis, Lucas; Huszár, Ferenc; Caballero, Jose; Cunningham, Andrew; Acosta, Alejandro; Aitken, Andrew; Tejani, Alykhan; Totz, Johannes; Wang, Zehan, and Shi, Wenzhe. Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 105–114, 2017. doi: 10.1109/CVPR.2017.19. (Cited on page 87.)

Lee, Dongha; Yu, Sehun; Ju, Hyunjun, and Yu, Hwanjo. Weakly supervised temporal anomaly segmentation with dynamic time warping. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7355–7364, 2021. (Cited on pages 44 and 113.)

Lee, Yunseung and Kang, Pilsung. Anovit: Unsupervised anomaly detection and localization with vision transformer-based encoder-decoder. *IEEE Access*, 10:46717–46724, 2022. (Cited on page 44.)

Lees, Thomas; Tseng, Gabriel; Atzberger, Clement; Reece, Steven, and Dadson, Simon. Deep learning for vegetation health forecasting: A case study in kenya. *Remote Sensing*, 14(3), 2022. ISSN 2072-4292. doi: 10.3390/rs14030698. URL https://www.mdpi.com/2072-4292/14/3/698. (Cited on page 46.)

Lehner, Bernhard and Grill, Günther. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes*, 27(15): 2171–2186, 2013. doi: https://doi.org/10.1002/hyp.9740. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.9740. (Cited on pages 35 and 224.)

Lessig, Christian; Luise, Ilaria; Gong, Bing; Langguth, Michael; Stadler, Scarlet, and Schultz, Martin. Atmorep: A stochastic model of atmosphere dynamics using large scale representation learning. *arXiv preprint arXiv:2308.13280*, 2023. (Cited on page 134.)

Li, Chun-Liang; Sohn, Kihyuk; Yoon, Jinsung, and Pfister, Tomas. Cutpaste: Self-supervised learning for anomaly detection and localization. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9659–9669, 2021. doi: 10.1109/CVPR46437.2021.00954. (Cited on page 44.)

Li, Jun; Geer, Alan J; Okamoto, Kozo; Otkin, Jason A; Liu, Zhiquan; Han, Wei, and Wang, Pei. Satellite all-sky infrared radiance assimilation: Recent progress and future perspectives. *Advances in Atmospheric Sciences*, 39(1):9–21, 2022a. doi: 10.1007/s00376-021-1088-9. (Cited on page 80.)

Li, Kunchang; Li, Xinhao; Wang, Yi; He, Yinan; Wang, Yali; Wang, Limin, and Qiao, Yu. Videomamba: State space model for efficient video understanding. In *European Conference on Computer Vision (ECCV) 2024*, pages 237–255, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-73347-5. doi: https://doi.org/10.1007/978-3-031-73347-5_14. (Cited on page 50.)

Li, Rui; Zheng, Shunyi; Duan, Chenxi; Yang, Yang, and Wang, Xiqi. Classification of hyperspectral image based on double-branch dual-attention mechanism network. *Remote Sensing*, 12(3), 2020. ISSN 2072-4292. doi: 10.3390/rs12030582. URL https://www.mdpi.com/2072-4292/12/3/582. (Cited on page 49.)

Li, Shijie; Cheng, Ming-Ming, and Gall, Juergen. Dual pyramid generative adversarial networks for semantic image synthesis. In *British Machine Vision Conference*, 2022b. (Cited on page 48.)

Liang, Dingkang; Zhou, Xin; Xu, Wei; Zhu, Xingkui; Zou, Zhikang; Ye, Xiaoqing; Tan, Xiao, and Bai, Xiang. Pointmamba: A simple state space model for point cloud analysis. In *Advances in Neural Information Processing Systems*, volume 37, pages 32653–32677. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/395371f778ebd4854b88521100af30ad-Paper-Conference.pdf. (Cited on page 50.)

Lin, Min; Chen, Qiang, and Yan, Shuicheng. Network in network. *arXiv preprint arXiv:1312.4400*, 2013. (Cited on page 54.)

Ling, Jun; Xue, Han; Song, Li; Xie, Rong, and Gu, Xiao. Region-aware adaptive instance normalization for image harmonization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9357–9366, 2021. doi: 10.1109/CVPR46437.2021.00924. (Cited on page 48.)

Linke, Simon; Lehner, Bernhard; Ouellet Dallaire, Camille; Ariwi, Joseph; Grill, Günther; Anand, Mira; Beames, Penny; Burchard-Levine, Vicente; Maxwell, Sally; Moidu, Hana, and others, . Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution. *Scientific data*, 6(1):283, 2019. doi: https://doi.org/10.1038/s41597-019-0300-6. (Cited on pages 35, 192 and 224.)

Linnainmaa, Seppo. Taylor expansion of the accumulated rounding error. *BIT Numerical Mathematics*, 16(2):146–160, 1976. doi: 10.1007/BF01931367. (Cited on page 12.)

Liu, Fei Tony; Ting, Kai Ming, and Zhou, Zhi-Hua. Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(1):1–39, 2012. (Cited on pages 117, 128, 130 and 131.)

Liu, Wenrui; Chang, Hong; Ma, Bingpeng; Shan, Shiguang, and Chen, Xilin. Diversity-measurable anomaly detection. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (CVPR)*, pages 12147–12156, 2023a. (Cited on page 44.)

Liu, Xianfeng; Zhu, Xiufang; Pan, Yaozhong; Li, Shuangshuang; Liu, Yanxu, and Ma, Yuqi. Agricultural drought monitoring: Progress, challenges, and prospects. *Journal of Geographical Sciences*, 26:750–767, 2016a. (Cited on page 24.)

Liu, Xun; Deng, Chenwei; Chanussot, Jocelyn; Hong, Danfeng, and Zhao, Baojun. Stfnet: A two-stream convolutional neural network for spatiotemporal image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6552–6564, 2019. doi: 10.1109/TGRS.2019.2907310. (Cited on page 49.)

Liu, Yue; Tian, Yunjie; Zhao, Yuzhong; Yu, Hongtian; Xie, Lingxi; Wang, Yaowei; Ye, Qixiang; Jiao, Jianbin, and Liu, Yunfan. Vmamba: Visual state space model. In *Advances in Neural Information Processing Systems*, volume 37, pages 103031–103063. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/baa2da9ae4bfed26520bb61d259a3653-Paper-Conference.pdf. (Cited on pages 50 and 158.)

Liu, Yunjie; Racah, Evan; Correa, Joaquin; Khosrowshahi, Amir; Lavers, David; Kunkel, Kenneth; Wehner, Michael; Collins, William, and others, . Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *arXiv preprint arXiv:1605.01156*, 2016b. (Cited on page 43.)

Liu, Ze; Lin, Yutong; Cao, Yue; Hu, Han; Wei, Yixuan; Zhang, Zheng; Lin, Stephen, and Guo, Baining. Swin transformer: Hierarchical vision transformer using shifted windows. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9992–10002, 2021. doi: 10.1109/ICCV48922.2021.00986. (Cited on pages 16, 17 and 88.)

Liu, Ze; Hu, Han; Lin, Yutong; Yao, Zhuliang; Xie, Zhenda; Wei, Yixuan; Ning, Jia; Cao, Yue; Zhang, Zheng; Dong, Li; Wei, Furu, and Guo, Baining. Swin transformer v2: Scaling up capacity and resolution. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11999–12009, 2022a. doi: 10.1109/CVPR52688.2022.01170. (Cited on pages 16 and 88.)

Liu, Ze; Ning, Jia; Cao, Yue; Wei, Yixuan; Zhang, Zheng; Lin, Stephen, and Hu, Han. Video swin transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3202–3211, June 2022b. (Cited on pages 17, 60, 61, 62, 63, 69, 111, 128, 130, 131 and 134.)

Liu, Zhikang; Zhou, Yiming; Xu, Yuansheng, and Wang, Zilei. Simplenet: A simple network for image anomaly detection and localization. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20402–20411, 2023b. doi: 10.1109/CVPR52729.2023.01954. (Cited on pages 44, 117, 128, 130 and 131.)

Liznerski, Philipp; Ruff, Lukas; Vandermeulen, Robert A.; Franks, Billy Joe; Kloft, Marius, and Müller, Klaus-Robert. Explainable deep one-class classification. In *International Conference on Learning Representations (ICLR)*, 2021. URL https://openreview.net/forum?id=A5VV3UyIQz. (Cited on page 44.)

Lorenz, Edward N. Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, 20(2):130 – 141, 1963. doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2. URL https://journals.ametsoc.org/view/journals/atsc/20/2/1520-0469_1963_020_0130_dnf_2_0_co_2.xml. (Cited on page 2.)

Lorenz, Edward N. *The Essence of Chaos*. University of Washington Press, 1993. (Cited on page 2.)

Loshchilov, Ilya and Hutter, Frank. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7. (Cited on page 89.)

Luo, Dezhao; Liu, Chang; Zhou, Yu; Yang, Dongbao; Ma, Can; Ye, Qixiang, and Wang, Weiping. Video cloze procedure for self-supervised spatio-temporal learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 11701–11708, 2020. (Cited on page 44.)

Lütjens, Björn; Leshchinskiy, Brandon; Boulais, Océane; Chishtie, Farrukh; Díaz-Rodríguez, Natalia; Masson-Forsythe, Margaux; Mata-Payerro, Ana; Requena-Mesa, Christian; Sankaranarayanan, Aruna; Piña, Aaron; Gal, Yarin; Raïssi, Chedy; Lavin, Alexander, and Newman, Dava. Generating physically-consistent satellite imagery for climate visualizations. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–11, 2024. doi: 10.1109/TGRS.2024.3493763. (Cited on page 47.)

Lv, Hui; Zhou, Chuanwei; Cui, Zhen; Xu, Chunyan; Li, Yong, and Yang, Jian. Localizing anomalies from weakly-labeled videos. *IEEE Transactions on Image Processing*, 30:4505–4515, 2021a. doi: 10.1109/TIP.2021.3072863. (Cited on page 45.)

Lv, Hui; Yue, Zhongqi; Sun, Qianru; Luo, Bin; Cui, Zhen, and Zhang, Hanwang. Unbiased multiple instance learning for weakly supervised video anomaly detection. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (CVPR)*, pages 8022–8031, 2023. (Cited on page 45.)

Lv, Zhengyao; Li, Xiaoming; Li, Xin; Li, Fu; Lin, Tianwei; He, Dongliang, and Zuo, Wangmeng. Learning semantic person image generation by region-adaptive normalization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10801–10810, 2021b. doi: 10.1109/CVPR46437.2021.01066. (Cited on page 48.)

Lyu, Yueming; Chen, Peibin; Sun, Jingna; Wang, Xu; Dong, Jing, and Tan, Tieniu. Detailed region-adaptive normalization for heavy makeup transfer. *arXiv preprint arXiv:2109.14525*, 2021. (Cited on page 48.)

Ma, Yueling; Montzka, Carsten; Bayat, Bagher, and Kollet, Stefan. An indirect approach based on long short-term memory networks to estimate groundwater table depth anomalies across europe with an application for drought analysis. *Frontiers in Water*, 3, 2021. ISSN 2624-9375. doi: 10.3389/frwa.2021.723548. URL https://www.frontiersin.org/articles/10.3389/frwa.2021.723548. (Cited on pages 79 and 103.)

Madan, Neelu; Ristea, Nicolae-Cătălin; Ionescu, Radu Tudor; Nasrollahi, Kamal; Khan, Fahad Shahbaz; Moeslund, Thomas B., and Shah, Mubarak. Self-supervised masked convolutional transformer block for anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(1):525–542, 2024. doi: 10.1109/TPAMI.2023.3322604. (Cited on page 44.)

Mardani, Morteza; Brenowitz, Noah; Cohen, Yair; Pathak, Jaideep; Chen, Chieh-Yu; Liu, Cheng-Chin; Vahdat, Arash; Nabian, Mohammad Amin; Ge, Tao; Subramaniam, Akshay, and others, . Residual corrective diffusion modeling for km-scale atmospheric downscaling. *Communications Earth & Environment*, 6(1):124, 2025. (Cited on page 219.)

Marín, Javier and Escalera, Sergio. Sssgan: Satellite style and structure generative adversarial networks. *Remote Sensing*, 13(19), 2021. ISSN 2072-4292. doi: 10.3390/rs13193984. URL https://www.mdpi.com/2072-4292/13/19/3984. (Cited on page 48.)

Marj, Ahmad Fatehi and Meijerink, Allard M. J. Agricultural drought forecasting using satellite images, climate indices and artificial neural network. *International Journal of Remote Sensing*, 32 (24):9707–9719, 2011. doi: 10.1080/01431161.2011.575896. URL https://doi.org/10.1080/01431161.2011.575896. (Cited on page 47.)

Martín, Yago; Zúñiga-Antón, María, and Mimbrero, Marcos Rodrigues. Modelling temporal variation of fire-occurrence towards the dynamic prediction of human wildfire ignition danger in northeast spain. *Geomatics, Natural Hazards and Risk*, 10(1):385–411, 2019. doi: 10.1080/19475705.2018.1526219. URL https://doi.org/10.1080/19475705.2018.1526219. (Cited on page 58.)

Martínez-Fernández, J.; González-Zamora, A.; Sánchez, N., and Gumuzzio, A. A soil water based index as a suitable agricultural drought indicator. *Journal of Hydrology*, 522:265–273, 2015. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2014.12.051. URL https://

www.sciencedirect.com/science/article/pii/S0022169414010622. (Cited on page 79.)

Maxwell, R. M.; Condon, L. E., and Kollet, S. J. A high-resolution simulation of groundwater and surface water over most of the continental us with the integrated hydrologic model parflow v3. *Geoscientific Model Development*, 8(3):923–937, 2015. doi: 10.5194/gmd-8-923-2015. URL https://gmd.copernicus.org/articles/8/923/2015/. (Cited on page 21.)

McKee, Thomas B. Drought monitoring with multiple time scales. In *Proceedings of 9th Conference on Applied Climatology, Boston, 1995*, 1995. (Cited on page 79.)

McNally, Anthony; Lessig, Christian; Lean, Peter; Boucher, Eulalie; Alexe, Mihai; Pinnington, Ewan; Chantry, Matthew; Lang, Simon; Burrows, Chris; Chrust, Marcin, and others, . Data driven weather forecasts trained and initialised directly from observations. *arXiv preprint arXiv:2407.15586*, 2024. (Cited on page 218.)

Mentzer, Fabian; Minnen, David; Agustsson, Eirikur, and Tschannen, Michael. Finite scalar quantization: VQ-VAE made simple. In *The Twelfth International Conference on Learning Representations (ICLR)*, 2024. URL https://openreview.net/forum?id=8ishA3LxN8. (Cited on page 138.)

Merz, Bruno; Blöschl, Günter; Vorogushyn, Sergiy; Dottori, Francesco; Aerts, Jeroen C. J. H.; Bates, Paul; Bertola, Miriam; Kemter, Matthias; Kreibich, Heidi; Lall, Upmanu, and others, . Causes, impacts and patterns of disastrous river floods. *Nature Reviews Earth & Environment*, 2(9):592–609, 2021. doi: https://doi.org/10.1038/s43017-021-00195-3. (Cited on page 154.)

Meza, I.; Siebert, S.; Döll, P.; Kusche, J.; Herbert, C.; Eyshi Rezaei, E.; Nouri, H.; Gerdener, H.; Popat, E.; Frischen, J.; Naumann, G.; Vogt, J. V.; Walz, Y.; Sebesvari, Z., and Hagenlocher, M. Global-scale drought risk assessment for agricultural systems. *Natural Hazards and Earth System Sciences*, 20(2):695–712, 2020. doi: 10.5194/nhess-20-695-2020. URL https://nhess.copernicus.org/articles/20/695/2020/. (Cited on pages 24 and 103.)

Miao, Lijuan; Ye, Peilong; He, Bin; Chen, Lizi, and Cui, Xuefeng. Future climate impact on the desertification in the dry land asia using avhrr gimms ndvi3g data. *Remote Sensing*, 7(4):3863–3877, 2015. ISSN 2072-4292. doi: 10.3390/rs70403863. URL https://www.mdpi.com/2072-4292/7/4/3863. (Cited on pages 47 and 104.)

Michail, Dimitrios; Davalas, Charalampos; Panagiotou, Lefki-Ioanna; Prapas, Ioannis; Kondylatos, Spyros; Bountos, Nikolaos Ioannis, and Papoutsis, Ioannis. Firecastnet: Earth-as-a-graph for seasonal fire prediction. *arXiv preprint arXiv:2502.01550*, 2025. (Cited on page 46.)

Miralles, Diego G.; Gentine, Pierre; Seneviratne, Sonia I., and Teuling, Adriaan J. Land–atmospheric feedbacks during droughts and heatwaves: state of the science and current challenges. *Annals of the New York Academy of Sciences*, 1436(1):19–35, 2019. doi: https://doi.org/10.1111/nyas.13912. URL https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/nyas.13912. (Cited on pages 79, 104, 108 and 140.)

Miranda, Miro; Zabawa, Laura; Kicherer, Anna; Strothmann, Laurenz; Rascher, Uwe, and Roscher, Ribana. Detection of anomalous grapevine berries using variational autoencoders. *Frontiers in Plant Science*, 13, 2022. ISSN 1664-462X. doi: 10.3389/fpls.2022.729097. URL https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2022.729097. (Cited on page 44.)

Mishra, Pankaj; Verk, Riccardo; Fornasier, Daniele; Piciarelli, Claudio, and Foresti, Gian Luca. VT-ADL: A vision transformer network for image anomaly detection and localization. In *30th IEEE/IES International Symposium on Industrial Electronics (ISIE)*, June 2021. (Cited on page 44.)

Mitsopoulos, Ioannis and Mallinis, Giorgos. A data-driven approach to assess large fire size generation in greece. *Natural Hazards*, 88:1591–1607, 2017. (Cited on page 45.)

Moravec, David; Komárek, Jan; López-Cuervo Medina, Serafín, and Molina, Iñigo. Effect of atmospheric corrections on ndvi: Intercomparability of landsat 8, sentinel-2, and uav sensors. *Remote Sensing*, 13(18), 2021. ISSN 2072-4292. doi: 10.3390/rs13183550. URL https://www.mdpi.com/2072-4292/13/18/3550. (Cited on page 93.)

Moreira, Francisco; Ascoli, Davide; Safford, Hugh; Adams, Mark A; Moreno, José M; Pereira, José M C; Catry, Filipe X; Armesto, Juan; Bond, William; González, Mauro E; Curt, Thomas; Koutsias, Nikos; McCaw, Lachlan; Price, Owen; Pausas, Juli G; Rigolot, Eric; Stephens, Scott; Tavsanoglu, Cagatay; Vallejo, V Ramon; Wilgen, Brian W Van; Xanthopoulos, Gavriil, and Fernandes, Paulo M. Wildfire management in mediterranean-type regions: paradigm change needed. *Environmental Research Letters*, 15(1):011001, jan 2020. doi: 10.1088/1748-9326/ab541e. URL https://dx.doi.org/10.1088/1748-9326/ab541e. (Cited on page 52.)

Mosavi, Amir; Ozturk, Pinar, and Chau, Kwok-wing. Flood prediction using machine learning models: Literature review. *Water*, 10(11):1536, 2018. (Cited on page 47.)

Moskolaï, Waytehad Rose; Abdou, Wahabou; Dipanda, Albert, and Kolyang, . Application of deep learning architectures for satellite image time series prediction: A review. *Remote Sensing*, 13 (23), 2021. ISSN 2072-4292. doi: 10.3390/rs13234822. URL https://www.mdpi.com/2072-4292/13/23/4822. (Cited on page 50.)

Muñoz Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; Martens, B.; Miralles, D. G.; Piles, M.; Rodríguez-Fernández, N. J.; Zsoter, E.; Buontempo, C., and Thépaut, J.-N. Era5-land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9):4349–4383, 2021. doi: 10.5194/essd-13-4349-2021. URL https://essd.copernicus.org/articles/13/4349/2021/. (Cited on pages 3, 20, 27, 29, 30, 223 and 224.)

Naderi, Mohammad Reza; Givkashi, Mohammad Hossein; Piri, Fatemeh; Karimi, Nader, and Samavi, Shadrokh. Focal-unet: Unet-like focal modulation for medical image segmentation. *arXiv preprint arXiv:2212.09263*, 2022. (Cited on page 18.)

Nagol, Jyoteshwar R.; Vermote, Eric F., and Prince, Stephen D. Effects of atmospheric variation on avhrr ndvi data. *Remote Sensing of Environment*, 113(2):392–397, 2009. ISSN 0034-4257.

doi: 10.1016/j.rse.2008.10.007. URL https://www.sciencedirect.com/science/article/pii/S0034425708003210. (Cited on pages 93 and 104.)

Najafi, Husain; Shrestha, Pallav Kumar; Rakovec, Oldrich; Apel, Heiko; Vorogushyn, Sergiy; Kumar, Rohini; Thober, Stephan; Merz, Bruno, and Samaniego, Luis. High-resolution impact-based early warning system for riverine flooding. *Nature communications*, 15(1):3726, 2024. doi: https://doi.org/10.1038/s41467-024-48065-y. (Cited on pages 155 and 198.)

Nay, John; Burchfield, Emily, and Gilligan, Jonathan. A machine-learning approach to forecasting remotely sensed vegetation health. *International Journal of Remote Sensing*, 39(6):1800–1816, 2018. doi: 10.1080/01431161.2017.1410296. URL https://doi.org/10.1080/01431161.2017.1410296. (Cited on page 46.)

Naz, B. S.; Sharples, W.; Ma, Y.; Goergen, K., and Kollet, S. Continental-scale evaluation of a fully distributed coupled land surface and groundwater model, parflow-clm (v3.6.0), over europe. *Geoscientific Model Development*, 16(6):1617–1639, 2023. doi: 10.5194/gmd-16-1617-2023. URL https://gmd.copernicus.org/articles/16/1617/2023/. (Cited on pages 79 and 93.)

Nearing, Grey. Global prediction of extreme floods in ungauged watersheds, December 2023. URL https://doi.org/10.5281/zenodo.10397664. (Cited on pages 192 and 195.)

Nearing, Grey; Cohen, Deborah; Dube, Vusumuzi; Gauch, Martin; Gilon, Oren; Harrigan, Shaun; Hassidim, Avinatan; Klotz, Daniel; Kratzert, Frederik; Metzger, Asher, and others, . Global prediction of extreme floods in ungauged watersheds. *Nature*, 627(8004):559–563, 2024. doi: https://doi.org/10.1038/s41586-024-07145-1. (Cited on pages 2, 32, 33, 48, 155, 158, 167, 169, 170, 171, 192, 193, 195, 208 and 218.)

Nearing, Grey S.; Kratzert, Frederik; Sampson, Alden Keefe; Pelissier, Craig S.; Klotz, Daniel; Frame, Jonathan M.; Prieto, Cristina, and Gupta, Hoshin V. What role does hydrological science play in the age of machine learning? *Water Resources Research*, 57(3):e2020WR028091, 2021. doi: https://doi.org/10.1029/2020WR028091. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020WR028091. e2020WR028091 10.1029/2020WR028091. (Cited on page 155.)

Nevo, S.; Morin, E.; Gerzi Rosenthal, A.; Metzger, A.; Barshai, C.; Weitzner, D.; Voloshin, D.; Kratzert, F.; Elidan, G.; Dror, G.; Begelman, G.; Nearing, G.; Shalev, G.; Noga, H.; Shavitt, I.; Yuklea, L.; Royz, M.; Giladi, N.; Peled Levi, N.; Reich, O.; Gilon, O.; Maor, R.; Timnat, S.; Shechter, T.; Anisimov, V.; Gigi, Y.; Levin, Y.; Moshe, Z.; Ben-Haim, Z.; Hassidim, A., and Matias, Y. Flood forecasting with machine learning models in an operational framework. *Hydrology and Earth System Sciences*, 26(15):4013–4032, 2022. doi: 10.5194/hess-26-4013-2022. URL https://hess.copernicus.org/articles/26/4013/2022/. (Cited on page 155.)

Nguyen, Tung; Brandstetter, Johannes; Kapoor, Ashish; Gupta, Jayesh K, and Grover, Aditya. Climax: A foundation model for weather and climate. *arXiv preprint arXiv:2301.10343*, 2023. (Cited on page 134.)

Nyborg, Joachim; Pelletier, Charlotte, and Assent, Ira. Generalized classification of satellite image time series with thermal positional encoding. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1391–1401, 2022a. doi: 10.1109/CVPRW56347.2022.00145. (Cited on page 50.)

Nyborg, Joachim; Pelletier, Charlotte; Lefèvre, Sébastien, and Assent, Ira. Timematch: Unsupervised cross-region adaptation by temporal shift estimation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 188:301–313, 2022b. ISSN 0924-2716. doi: https://doi.org/10.1016/j.isprsjprs.2022.04.018. URL https://www.sciencedirect.com/science/article/pii/S0924271622001216. (Cited on pages 50 and 74.)

Oleson, K. W.; Niu, G.-Y.; Yang, Z.-L.; Lawrence, D. M.; Thornton, P. E.; Lawrence, P. J.; Stöckli, R.; Dickinson, R. E.; Bonan, G. B.; Levis, S.; Dai, A., and Qian, T. Improvements to the community land model and their impact on the hydrological cycle. *Journal of Geophysical Research: Biogeosciences*, 113(G1), 2008. doi: 10.1029/2007JG000563. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JG000563. (Cited on page 21.)

Oleson, Keith; YJ, Dai; Bosilovich, M.; Dickinson, Robert; Dirmeyer, Paul; Hoffman, Forrest; Houser, P.; Levis, Samuel; Niu, Guo-Yue; Thornton, P.; Vertenstein, Mariana; Yang, Zong-Liang, and XB, Zeng. *Technical Description of the Community Land Model (CLM)*, page 174 pp. 05 2004. ISBN NCAR/TN-461+STR. doi: 10.5065/D6N877R0. (Cited on page 21.)

Ou, Zhigang; Nai, Congyi; Pan, Baoxiang; Zheng, Yi; Shen, Chaopeng; Jiang, Peishi; Liu, Xingcai; Tang, Qiuhong; Li, Wenqing, and Pan, Ming. Probabilistic diffusion models advance extreme flood forecasting. *Geophysical Research Letters*, 52(15):e2025GL115705, 2025. doi: https://doi.org/10.1029/2025GL115705. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2025GL115705. e2025GL115705 2025GL115705. (Cited on pages 47 and 215.)

Palash, Wahid; Akanda, Ali S, and Islam, Shafiqul. A data-driven global flood forecasting system for medium to large rivers. *Scientific Reports*, 14(1):8979, 2024. doi: https://doi.org/10.1038/s41598-024-59145-w. (Cited on page 47.)

Pappenberger, Florian; Cloke, Hannah L.; Parker, Dennis J.; Wetterhall, Fredrik; Richardson, David S., and Thielen, Jutta. The monetary benefit of early flood warnings in europe. *Environmental Science & Policy*, 51:278–291, 2015. ISSN 1462-9011. doi: https://doi.org/10.1016/j.envsci.2015.04.016. URL https://www.sciencedirect.com/science/article/pii/S1462901115000891. (Cited on page 155.)

Park, Seongheon; Kim, Hanjae; Kim, Minsu; Kim, Dahye, and Sohn, Kwanghoon. Normality guided multiple instance learning for weakly supervised video anomaly detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2665–2674, January 2023. (Cited on page 45.)

Park, Taesung; Liu, Ming-Yu; Wang, Ting-Chun, and Zhu, Jun-Yan. Semantic image synthesis with spatially-adaptive normalization. In *IEEE/CVF Conference on Computer Vision and Pat-*

tern Recognition (CVPR), pages 2332–2341, 2019. doi: 10.1109/CVPR.2019.00244. (Cited on page 48.)

Paszke, Adam; Gross, Sam; Massa, Francisco; Lerer, Adam; Bradbury, James; Chanan, Gregory; Killeen, Trevor; Lin, Zeming; Gimelshein, Natalia; Antiga, Luca; Desmaison, Alban; Kopf, Andreas; Yang, Edward; DeVito, Zachary; Raison, Martin; Tejani, Alykhan; Chilamkurthy, Sasank; Steiner, Benoit; Fang, Lu; Bai, Junjie, and Chintala, Soumith. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf. (Cited on pages 58 and 89.)

Patacca, Marco; Lindner, Marcus; Lucas-Borja, Manuel Esteban; Cordonnier, Thomas; Fidej, Gal; Gardiner, Barry; Hauf, Ylva; Jasinevičius, Gediminas; Labonne, Sophie; Linkevičius, Edgaras; Mahnken, Mats; Milanovic, Slobodan; Nabuurs, Gert-Jan; Nagel, Thomas A.; Nikinmaa, Laura; Panyatov, Momchil; Bercak, Roman; Seidl, Rupert; Ostrogović Sever, Masa Zorana; Socha, Jaroslaw; Thom, Dominik; Vuletic, Dijana; Zudin, Sergey, and Schelhaas, Mart-Jan. Significant increase in natural disturbance impacts on european forests since 1950. *Global Change Biology*, 29(5):1359–1376, 2023. doi: 10.1111/gcb.16531. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/gcb.16531. (Cited on page 79.)

Patakchi Yousefi, Kaveh and Kollet, Stefan. Deep learning of model- and reanalysis-based precipitation and pressure mismatches over europe. *Frontiers in Water*, 5, 2023. ISSN 2624-9375. doi: 10.3389/frwa.2023.1178114. URL https://www.frontiersin.org/articles/10.3389/frwa.2023.1178114. (Cited on page 79.)

Pathak, Jaideep; Subramanian, Shashank; Harrington, Peter; Raja, Sanjeev; Chattopadhyay, Ashesh; Mardani, Morteza; Kurth, Thorsten; Hall, David; Li, Zongyi; Azizzadenesheli, Kamyar, and others, . Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022. (Cited on pages 155 and 218.)

Patil, Sopan D.; Gu, Yuting; Dias, Felipe S. A; Stieglitz, Marc, and Turk, Greg. Predicting the spectral information of future land cover using machine learning. *International Journal of Remote Sensing*, 38(20):5592–5607, 2017. doi: 10.1080/01431161.2017.1343512. URL https://doi.org/10.1080/01431161.2017.1343512. (Cited on page 47.)

Peano, Giuseppe and Peano, G. *Sur une courbe, qui remplit toute une aire plane*. Springer, 1990. doi: https://doi.org/10.1007/BF01199438. (Cited on page 159.)

Pedro-Monzonís, María; Solera, Abel; Ferrer, Javier; Estrela, Teodoro, and Paredes-Arquiola, Javier. A review of water scarcity and drought indexes in water resources planning and management. *Journal of Hydrology*, 527:482–493, 2015. (Cited on page 24.)

Pei, Xiaohuan; Huang, Tao, and Xu, Chang. Efficientvmamba: Atrous selective scan for light weight visual mamba. *arXiv preprint arXiv:2403.09977*, 2024. (Cited on page 50.)

Pelletier, Charlotte; Webb, Geoffrey I, and Petitjean, François. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, 11(5):523, 2019. https://www.mdpi.com/2072-4292/11/5/523. (Cited on page 50.)

Perkins-Kirkpatrick, SE and Lewis, SC. Increasing trends in regional heatwaves. *Nature communications*, 11(1):1–8, 2020. (Cited on pages 1 and 52.)

Pham, Binh Thai; Jaafari, Abolfazl; Avand, Mohammadtaghi; Al-Ansari, Nadhir; Dinh Du, Tran; Yen, Hoang Phan Hai; Phong, Tran Van; Nguyen, Duy Huu; Le, Hiep Van; Mafi-Gholami, Davood; Prakash, Indra; Thi Thuy, Hoang, and Tuyen, Tran Thi. Performance evaluation of machine learning methods for forest fire modeling and prediction. *Symmetry*, 12(6), 2020. ISSN 2073-8994. doi: 10.3390/sym12061022. URL https://www.mdpi.com/2073-8994/12/6/1022. (Cited on page 45.)

Phung, Hao; Dao, Quan; Dao, Trung; Phan, Hoang; Metaxas, Dimitris N., and Tran, Anh. Dimsum: Diffusion mamba - a scalable and unified spatial-frequency method for image generation. In *Advances in Neural Information Processing Systems*, volume 37, pages 32947–32979. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/39bc6e3cbf5a1991d33dc10ebff9a9cf-Paper-Conference.pdf. (Cited on page 50.)

Pianforini, Matteo; Dazzi, Susanna; Pilzer, Andrea, and Vacondio, Renato. Floodsformer: A transformer-based data-driven model for predicting the 2-d dynamics of fluvial floods. *Environmental Modelling & Software*, 193:106599, 2025. ISSN 1364-8152. doi: https://doi.org/10.1016/j.envsoft.2025.106599. URL https://www.sciencedirect.com/science/article/pii/S136481522500283X. (Cited on page 47.)

Pirret, J.S.R.; Fung, F.; Lowe, J.A.; McInnes, R.N.; Mitchell, J.F.B., and Murphy, J.M. Ukcp factsheet: Soil moisture, 2020. URL https://www.metoffice.gov.uk/research/approach/collaboration/ukcp/data/factsheets. (Cited on page 79.)

Pokhrel, Yadu; Felfelani, Farshid; Satoh, Yusuke; Boulange, Julien Eric Stanislas; Burek, Peter; Gädeke, Anne; Gerten, Dieter; Gosling, Simon Newland; Grillakis, Manolis G.; Gudmundsson, Lukas; Hanasaki, Naota; Kim, Hyungjun; Koutroulis, Aristeidis G.; Liu, Junguo; Papadimitriou, Lamprini; Schewe, Jacob; Schmied, Hannes Müller; Stacke, Tobias Ingo; Telteu, Camelia-Eliza; Thiery, Wim; Veldkamp, Ted Isis Elize; Zhao, Fang, and Wada, Yoshihide. Global terrestrial water storage and drought severity under climate change. *Nature Climate Change*, 11:226 – 233, 2021. doi: 10.1038/s41558-020-00972-w. (Cited on page 79.)

Poshyvailo-Strube, L.; Wagner, N.; Goergen, K.; Furusho-Percot, C.; Hartick, C., and Kollet, S. Groundwater in terrestrial systems modelling: a new climatology of extreme heat events in europe. *Earth System Dynamics Discussions*, 2022:1–25, 2022. doi: 10.5194/esd-2022-53. URL https://esd.copernicus.org/preprints/esd-2022-53/. (Cited on page 22.)

Pourreza, Masoud; Mohammadi, Bahram; Khaki, Mostafa; Bouindour, Samir; Snoussi, Hichem, and Sabokrou, Mohammad. G2d: Generate to detect anomaly. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2003–2012, 2021. (Cited on page 44.)

Prabhat, ; Kashinath, K.; Mudigonda, M.; Kim, S.; Kapp-Schwoerer, L.; Graubner, A.; Karaismailoglu, E.; von Kleist, L.; Kurth, T.; Greiner, A.; Mahesh, A.; Yang, K.; Lewis, C.; Chen, J.; Lou, A.; Chandran, S.; Toms, B.; Chapman, W.; Dagon, K.; Shields, C. A.; O'Brien, T.; Wehner, M., and Collins, W. Climatenet: an expert-labeled open dataset and deep learning architecture for enabling high-precision analyses of extreme weather. *Geoscientific Model Development*, 14(1): 107–124, 2021. doi: 10.5194/gmd-14-107-2021. URL https://gmd.copernicus.org/articles/14/107/2021/. (Cited on page 43.)

Prapas, Ioannis; Kondylatos, Spyros; Papoutsis, Ioannis; Camps-Valls, Gustau; Ronco, Michele; Fernández-Torres, Miguel-Ángel; Guillem, Maria Piles, and Carvalhais, Nuno. Deep learning methods for daily wildfire danger forecasting, 2021. URL https://arxiv.org/abs/2111.02736. (Cited on pages 19, 45, 46, 53 and 60.)

Prapas, Ioannis; Ahuja, Akanksha; Kondylatos, Spyros; Karasante, Ilektra; Alonso, Lazaro; Panagiotou, Eleanna; Davalas, Charalampos; Michail, Dimitrios; Carvalhais, Nuno, and Papoutsis, Ioannis. Deep learning for global wildfire forecasting. In *NeurIPS Workshop on Tackling Climate Change with Machine Learning*, 2022a. URL https://www.climatechange.ai/papers/neurips2022/52. (Cited on page 45.)

Prapas, Ioannis; Kondylatos, Spyros, and Papoutsis, Ioannis. FireCube: A Daily Datacube for the Modeling and Analysis of Wildfires in Greece, May 2022b. URL https://doi.org/10.5281/zenodo.6475592. (Cited on page 19.)

Prapas, Ioannis; Kondylatos, Spyros, and Papoutsis, Ioannis. Training data for submitted paper "Wildfire Danger Prediction and Understanding with Deep Learning", May 2022c. URL https://doi.org/10.5281/zenodo.6528394. (Cited on pages 20 and 223.)

Prapas, Ioannis; Bountos, Nikolaos-Ioannis; Kondylatos, Spyros; Michail, Dimitrios; Camps-Valls, Gustau, and Papoutsis, Ioannis. Televit: Teleconnection-driven transformers improve subseasonal to seasonal wildfire forecasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 3754–3759, October 2023. (Cited on page 46.)

Prein, Andreas F; Rasmussen, Roy M; Ikeda, Kyoko; Liu, Changhai; Clark, Martyn P, and Holland, Greg J. The future intensification of hourly precipitation extremes. *Nature climate change*, 7(1): 48–52, 2017. doi: 10.1038/nclimate3168. (Cited on page 1.)

Price, Ilan; Sanchez-Gonzalez, Alvaro; Alet, Ferran; Andersson, Tom R; El-Kadi, Andrew; Masters, Dominic; Ewalds, Timo; Stott, Jacklynn; Mohamed, Shakir; Battaglia, Peter, and others, . Probabilistic weather forecasting with machine learning. *Nature*, 637(8044):84–90, 2025. doi: https://doi.org/10.1038/s41586-024-08252-9. (Cited on pages 155, 215 and 218.)

Prodhan, Foyez Ahmed; Zhang, Jiahua; Yao, Fengmei; Shi, Lamei; Pangali Sharma, Til Prasad; Zhang, Da; Cao, Dan; Zheng, Minxuan; Ahmed, Naveed, and Mohana, Hasiba Pervin. Deep learning for monitoring agricultural drought in south asia using remote sensing data. *Remote Sensing*, 13(9), 2021. ISSN 2072-4292. doi: 10.3390/rs13091715. URL https://www.mdpi.com/2072-4292/13/9/1715. (Cited on page 46.)

Qi, Zhenting; Zhu, Ruike; Fu, Zheyu; Chai, Wenhao, and Kindratenko, Volodymyr. Weakly supervised two-stage training scheme for deep video fight detection model. In *2022 IEEE 34th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 677–685. IEEE, 2022. (Cited on page 45.)

Qin, Qiming; Wu, Zihua; Zhang, Tianyuan; Sagan, Vasit; Zhang, Zhaoxu; Zhang, Yao; Zhang, Chengye; Ren, Huazhong; Sun, Yuanheng; Xu, Wei, and Zhao, Cong. Optical and thermal remote sensing for monitoring agricultural drought. *Remote Sensing*, 13(24), 2021. ISSN 2072-4292. doi: 10.3390/rs13245092. URL https://www.mdpi.com/2072-4292/13/24/5092. (Cited on page 79.)

Racah, Evan; Beckham, Christopher; Maharaj, Tegan; Ebrahimi Kahou, Samira; Prabhat, Mr, and Pal, Chris. Extremeweather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. *Advances in neural information processing systems*, 30, 2017. (Cited on page 43.)

Rasoulian, Amirhossein; Salari, Soorena, and Xiao, Yiming. Weakly supervised segmentation of intracranial aneurysms using a 3d focal modulation unet. *arXiv preprint arXiv:2308.03001*, 2023. (Cited on page 18.)

Reichstein, Markus; Camps-Valls, Gustau; Stevens, Bjorn; Jung, Martin; Denzler, Joachim; Carvalhais, Nuno, and others, . Deep learning and process understanding for data-driven earth system science. *Nature*, 566(7743):195–204, 2019. (Cited on pages 2 and 52.)

Reichstein, Markus; Benson, Vitus; Blunk, Jan; Camps-Valls, Gustau; Creutzig, Felix; Fearnley, Carina J; Han, Boran; Kornhuber, Kai; Rahaman, Nasim; Schölkopf, Bernhard, and others, . Early warning of complex climate risk with integrated artificial intelligence. *Nature Communications*, 16(1):2564, 2025. doi: https://doi.org/10.1038/s41467-025-57640-w. (Cited on pages 2 and 155.)

Reiss, Tal; Cohen, Niv; Bergman, Liron, and Hoshen, Yedid. Panda: Adapting pretrained features for anomaly detection and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2806–2814, June 2021. (Cited on page 44.)

Ren, Li; Arkin, Phillip; Smith, Thomas M., and Shen, Samuel S.P. Global precipitation trends in 1900–2005 from a reconstruction and coupled model simulations. *Journal of Geophysical Research: Atmospheres*, 118(4):1679–1689, 2013. doi: https://doi.org/10.1002/jgrd.50212. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jgrd.50212. (Cited on page 52.)

Ren, Shaoqing; He, Kaiming; Girshick, Ross, and Sun, Jian. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. (Cited on page 14.)

Rentschler, Jun; Salhab, Melda, and Jafino, Bramka Arga. Flood exposure and poverty in 188 countries. *Nature communications*, 13(1):3527, 2022. doi: https://doi.org/10.1038/s41467-022-30727-4. (Cited on page 155.)

Requena-Mesa, Christian; Benson, Vitus; Reichstein, Markus; Runge, Jakob, and Denzler, Joachim. Earthnet2021: A large-scale dataset and challenge for earth surface forecasting as a guided video prediction task. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1132–1142, 2021. doi: 10.1109/CVPRW53098.2021.00124. (Cited on page 46.)

Reyniers, N.; Osborn, T. J.; Addor, N., and Darch, G. Projected changes in droughts and extreme droughts in great britain strongly influenced by the choice of drought index. *Hydrology and Earth System Sciences*, 27(5):1151–1171, 2023. doi: 10.5194/hess-27-1151-2023. URL https://hess.copernicus.org/articles/27/1151/2023/. (Cited on page 79.)

Rippel, Oliver; Mertens, Patrick; König, Eike, and Merhof, Dorit. Gaussian anomaly detection by modeling the distribution of normal data in pretrained deep features. *IEEE Transactions on Instrumentation and Measurement*, 70:1–13, 2021. doi: 10.1109/TIM.2021.3098381. (Cited on page 44.)

Robin, Claire; Requena-Mesa, Christian; Benson, Vitus; Poehls, Jeran; Alonzo, Lazaro; Carvalhais, Nuno, and Reichstein, Markus. Learning to forecast vegetation greenness at fine resolution over africa with convlstms. In *NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning*, 2022. URL https://www.climatechange.ai/papers/neurips2022/88. (Cited on page 46.)

Ronneberger, Olaf; Fischer, Philipp, and Brox, Thomas. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. doi: 10.1007/978-3-319-24574-4_28. (Cited on pages 81 and 88.)

Rosenblatt, Frank. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958. (Cited on page 12.)

Roth, Karsten; Pemula, Latha; Zepeda, Joaquin; Schölkopf, Bernhard; Brox, Thomas, and Gehler, Peter. Towards total recall in industrial anomaly detection. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14298–14308, 2022. doi: 10.1109/CVPR52688.2022.01392. (Cited on page 44.)

Roudbari, Naghmeh Shafiee; Punekar, Shubham Rajeev; Patterson, Zachary; Eicker, Ursula, and Poullis, Charalambos. From data to action in flood forecasting leveraging graph neural networks and digital twin visualization. *Scientific reports*, 14(1):18571, 2024. doi: https://doi.org/10.1038/s41598-024-68857-y. (Cited on page 47.)

Rumelhart, David E; Hinton, Geoffrey E, and Williams, Ronald J. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986. doi: 10.1038/323533a0. (Cited on page 12.)

Runge, Jakob; Bathiany, Sebastian; Bollt, Erik; Camps-Valls, Gustau; Coumou, Dim; Deyle, Ethan; Glymour, Clark; Kretschmer, Marlene; Mahecha, Miguel D; Muñoz-Marí, Jordi, and others, .

Inferring causation from time series in earth system sciences. *Nature communications*, 10(1): 2553, 2019a. doi: 10.1038/s41467-019-10105-3. (Cited on page 3.)

Runge, Jakob; Nowack, Peer; Kretschmer, Marlene; Flaxman, Seth, and Sejdinovic, Dino. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5 (11):eaau4996, 2019b. doi: 10.1126/sciadv.aau4996. URL https://www.science.org/doi/abs/10.1126/sciadv.aau4996. (Cited on page 3.)

Runge, Jakob; Gerhardus, Andreas; Varando, Gherardo; Eyring, Veronika, and Camps-Valls, Gustau. Causal inference for time series. *Nature Reviews Earth & Environment*, 4(7):487–505, 2023. doi: 10.1038/s43017-023-00431-y. (Cited on page 3.)

Ruparell, Karan; Marks, Robert J.; Wood, Andy; Hunt, Kieran M. R.; Cloke, Hannah L.; Prudhomme, Christel; Pappenberger, Florian, and Chantry, Matthew. Hydra-lstm: A semi-shared machine learning architecture for prediction across watersheds. *Artificial Intelligence for the Earth Systems*, 4(3):240103, 2025. doi: 10.1175/AIES-D-24-0103.1. URL https://journals.ametsoc.org/view/journals/aies/4/3/AIES-D-24-0103.1.xml. (Cited on page 47.)

Rußwurm, Marc and Körner, Marco. Multi-temporal land cover classification with sequential recurrent encoders. *ISPRS International Journal of Geo-Information*, 7(4), 2018. ISSN 2220-9964. doi: 10.3390/ijgi7040129. URL https://www.mdpi.com/2220-9964/7/4/129. (Cited on page 50.)

Sabokrou, Mohammad; Pourreza, Masoud; Fayyaz, Mohsen; Entezari, Rahim; Fathy, Mahmood; Gall, Jürgen, and Adeli, Ehsan. Avid: Adversarial visual irregularity detection. In *Computer Vision – ACCV 2018*, pages 488–505, Cham, 2019. Springer International Publishing. ISBN 978-3-030-20876-9. (Cited on page 44.)

Salakpi, E. E.; Hurley, P. D.; Muthoka, J. M.; Barrett, A. B.; Bowell, A.; Oliver, S., and Rowhani, P. Forecasting vegetation condition with a bayesian auto-regressive distributed lags (bardl) model. *Natural Hazards and Earth System Sciences*, 22(8):2703–2723, 2022a. doi: 10.5194/nhess-22-2703-2022. URL https://nhess.copernicus.org/articles/22/2703/2022/. (Cited on page 46.)

Salakpi, E. E.; Hurley, P. D.; Muthoka, J. M.; Bowell, A.; Oliver, S., and Rowhani, P. A dynamic hierarchical bayesian approach for forecasting vegetation condition. *Natural Hazards and Earth System Sciences*, 22(8):2725–2749, 2022b. doi: 10.5194/nhess-22-2725-2022. URL https://nhess.copernicus.org/articles/22/2725/2022/. (Cited on page 46.)

Salcedo-Sanz, Sancho; Pérez-Aracil, Jorge; Ascenso, Guido; Del Ser, Javier; Casillas-Pérez, David; Kadow, Christopher; Fister, Dušan; Barriopedro, David; García-Herrera, Ricardo; Giuliani, Matteo, and others, . Analysis, characterization, prediction, and attribution of extreme atmospheric events with machine learning and deep learning techniques: a review. *Theoretical and Applied Climatology*, 155(1):1–44, 2024. (Cited on page 43.)

Salehi, Mohammad reza; Sadjadi, Niousha; Baselizadeh, Soroosh; Rohban, Mohammad H., and Rabiee, Hamid R. Multiresolution knowledge distillation for anomaly detection. In *Proceedings*

of the IEEE/CVF Conference on computer vision and pattern recognition (CVPR), pages 14902–14912, 2021. (Cited on page 44.)

Samuels, Rana; Hochman, Assaf; Baharad, Anat; Givati, Amir; Levi, Yoav; Yosef, Yizhak; Saaroni, Hadas; Ziv, Baruch; Harpaz, Tzvika, and Alpert, Pinhas. Evaluation and projection of extreme precipitation indices in the eastern mediterranean based on cmip5 multi-model ensemble. *International Journal of Climatology*, 38(5):2280–2297, 2018. (Cited on page 52.)

San-Miguel-Ayanz, Jesús; Schulte, Ernst; Schmuck, Guido, and Camia, Andrea. The european forest fire information system in the context of environmental policies of the european union. *Forest Policy and Economics*, 29:19–25, 2013. ISSN 1389-9341. doi: https://doi.org/10.1016/j.forpol.2011.08.012. URL https://www.sciencedirect.com/science/article/pii/S138993411200127X. The FIRE PARADOX project: Setting the basis for a shift in the forest fire policies in Europe. (Cited on page 20.)

Sapkota, Hitesh and Yu, Qi. Bayesian nonparametric submodular video partition for robust anomaly detection. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (CVPR)*, pages 3212–3221, 2022. (Cited on page 45.)

Satoh, Yusuke; Shiogama, Hideo; Hanasaki, Naota; Pokhrel, Yadu; Boulange, Julien Eric Stanislas; Burek, Peter; Gosling, Simon Newland; Grillakis, Manolis; Koutroulis, Aristeidis; Schmied, Hannes Müller; Thiery, Wim, and Yokohata, Tokuta. A quantitative evaluation of the issue of drought definition: a source of disagreement in future drought assessments. *Environmental Research Letters*, 16(10):104001, sep 2021. doi: 10.1088/1748-9326/ac2348. URL https://dx.doi.org/10.1088/1748-9326/ac2348. (Cited on page 79.)

Scheck, Leonhard; Frèrebeau, Pascal; Buras-Schnell, Robert, and Mayer, Bernhard. A fast radiative transfer method for the simulation of visible satellite imagery. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 175:54–67, 2016. ISSN 0022-4073. doi: 10.1016/j.jqsrt.2016.02.008. URL https://www.sciencedirect.com/science/article/pii/S0022407316000285. (Cited on page 80.)

Schimanke, S.; Ridal, M.; Le Moigne, P.; Berggren, L.; Undén, P.; Randriamampianina, R.; Andrea, U.; Bazile, E.; Bertelsen, A.; Brousseau, P.; Dahlgren, P.; Edvinsson, L.; El Said, A.; Glinton, M.; Hopsch, S.; Isaksson, L.; Mladek, R.; Olsson, E.; Verrelle, A., and Wang, Z.Q. Cerra sub-daily regional reanalysis data for europe on single levels from 1984 to present, 2021. URL https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-cerra-single-levels?tab=overview. (Cited on pages 28, 31 and 223.)

Schlüter, Hannah M.; Tan, Jeremy; Hou, Benjamin, and Kainz, Bernhard. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *European Conference on Computer Vision (ECCV) 2022*, pages 474–489, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-19821-2. (Cited on page 44.)

Schmidhuber, Jürgen. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2014.09.003. URL https://

www.sciencedirect.com/science/article/pii/S0893608014002135. (Cited on page 12.)

Schölkopf, Bernhard; Williamson, Robert C; Smola, Alex; Shawe-Taylor, John, and Platt, John. Support vector method for novelty detection. *Advances in neural information processing systems*, 12, 1999. (Cited on pages 44, 117, 128, 130 and 131.)

Schölkopf, Bernhard; Locatello, Francesco; Bauer, Stefan; Ke, Nan Rosemary; Kalchbrenner, Nal; Goyal, Anirudh, and Bengio, Yoshua. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021. doi: 10.1109/JPROC.2021.3058954. (Cited on page 3.)

Schultz, M. G.; Betancourt, C.; Gong, B.; Kleinert, F.; Langguth, M.; Leufen, L. H.; Mozaffari, A., and Stadtler, S. Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194):20200097, 2021. doi: 10.1098/rsta.2020.0097. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2020.0097. (Cited on page 2.)

Scudo, Fabrizio Lo; De Rango, Alessio; Furnari, Luca; Senatore, Alfonso; D'Ambrosio, Donato; Mendicino, Giuseppe, and Greco, Gianluigi. Advancing wildfire risk prediction via morphology-aware curriculum contrastive learning. *arXiv preprint arXiv:2507.21147*, 2025. (Cited on page 62.)

Seleem, Omar; Ayzel, Georgy; de Souza, Arthur Costa Tomaz; Bronstert, Axel, and and, Maik Heistermann. Towards urban flood susceptibility mapping using data-driven models in berlin, germany. *Geomatics, Natural Hazards and Risk*, 13(1):1640–1662, 2022. doi: 10.1080/19475705.2022.2097131. URL https://doi.org/10.1080/19475705.2022.2097131. (Cited on page 47.)

Seneviratne, S.I.; Zhang, X.; Adnan, M.; Badi, W.; Dereczynski, C.; Di Luca, A.; Ghosh, S.; Iskandar, I.; Kossin, J.; Lewis, S.; Otto, F.; Pinto, I.; Satoh, M.; Vicente-Serrano, S.M.; Wehner, M., and Zhou, B. *Weather and Climate Extreme Events in a Changing Climate*, page 1513–1766. Cambridge University Press, 08 2021. doi: 10.1017/9781009157896.013. (Cited on pages 1 and 79.)

Sepulcre-Canto, G.; Horion, S.; Singleton, A.; Carrao, H., and Vogt, J. Development of a combined drought indicator to detect agricultural drought in europe. *Natural Hazards and Earth System Sciences*, 12(11):3519–3531, 2012. doi: 10.5194/nhess-12-3519-2012. URL https://nhess.copernicus.org/articles/12/3519/2012/. (Cited on pages 25 and 43.)

Shaker, Abdelrahman; Wasim, Syed Talal; Khan, Salman; Gall, Juergen, and Khan, Fahad Shahbaz. Groupmamba: Efficient group-based visual state space model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14912–14922, June 2025. (Cited on page 50.)

Shalev, Guy and Kratzert, Frederik. Caravan multimet: Extending caravan with multiple weather nowcasts and forecasts. *arXiv preprint arXiv:2411.09459*, 2024. (Cited on page 219.)

Shams Eddin, M. H. and Gall, J. Focal-tsmp: deep learning for vegetation health prediction and agricultural drought assessment from a regional climate simulation. *Geoscientific Model Development*, 17(7):2987–3023, 2024a. doi: 10.5194/gmd-17-2987-2024. URL https://gmd.copernicus.org/articles/17/2987/2024/. (Cited on pages 9 and 77.)

Shams Eddin, Mohamad Hakam and Gall, Juergen. Focal-tsmp: Deep learning for vegetation health prediction and agricultural drought assessment from a regional climate simulation. https://doi.org/10.5281/zenodo.10015048, October 2023a. (Cited on page 223.)

Shams Eddin, Mohamad Hakam and Gall, Juergen. Focal-tsmp: Deep learning for vegetation health prediction and agricultural drought assessment from a regional climate simulation. https://doi.org/10.5281/zenodo.10008814, October 2023b. (Cited on page 223.)

Shams Eddin, Mohamad Hakam and Gall, Juergen. Identifying spatio-temporal drivers of extreme events [data set], 2024b. URL https://doi.org/10.60507/FK2/RD9E33. (Cited on page 223.)

Shams Eddin, Mohamad Hakam and Gall, Jürgen. Identifying spatio-temporal drivers of extreme events. In *Advances in Neural Information Processing Systems*, volume 37, pages 93714–93766. Curran Associates, Inc., 2024c. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/aa7259c82d642e47d5661f3218cdcad2-Paper-Conference.pdf. (Cited on pages 9 and 107.)

Shams Eddin, Mohamad Hakam; Roscher, Ribana, and Gall, Juergen. Location-aware adaptive normalization: A deep learning approach for wildfire danger forecasting. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–18, 2023. doi: 10.1109/TGRS.2023.3285401. (Cited on pages 9 and 51.)

Shams Eddin, Mohamad Hakam; Zhang, Yikui; Kollet, Stefan, and Gall, Juergen. Rivermamba: A state space model for global river discharge and flood forecasting. *arXiv preprint arXiv:2505.22535*, 2025a. (Cited on pages 9 and 153.)

Shams Eddin, Mohamad Hakam; Zhang, Yikui; Kollet, Stefan, and Gall, Juergen. RiverMamba: A State Space Model for Global River Discharge and Flood Forecasting [data set], 2025b. URL https://doi.org/10.60507/FK2/T8QYWE. (Cited on page 224.)

Shang, Chen; Wulder, Michael A.; Coops, Nicholas C.; White, Joanne C., and Hermosilla, Txomin. Spatially-explicit prediction of wildfire burn probability using remotely-sensed and ancillary data. *Canadian Journal of Remote Sensing*, 46:313 – 329, 2020. (Cited on page 45.)

Sheffield, Justin; Wood, Eric F., and Roderick, Michael L. Little change in global drought over the past 60 years. *Nature*, 491:435–438, 2012. URL https://api.semanticscholar.org/CorpusID:205231154. (Cited on page 79.)

Shen, Yu; Xiao, Liang; Chen, Jianyu, and Pan, Delu. A spectral-spatial domain-specific convolutional deep extreme learning machine for supervised hyperspectral image classification. *IEEE Access*, 7:132240–132252, 2019. doi: 10.1109/ACCESS.2019.2940697. (Cited on page 49.)

Shi, Xiaokang; Li, Yaodong; Liu, Jianwen; Xiang, Xizi, and Liu, Le. Simulation of fy-2d infrared brightness temperature and sensitivity analysis to the errors of wrf simulated cloud variables. *Science China Earth Sciences*, 61:1–16, 03 2018. doi: 10.1007/s11430-017-9150-0. (Cited on page 104.)

Shi, Xingjian; Chen, Zhourong; Wang, Hao; Yeung, Dit-Yan; Wong, Wai-kin, and Woo, Wang-chun. Convolutional lstm network: a machine learning approach for precipitation nowcasting. In *Proceedings of the 28th International Conference on Neural Information Processing Systems*, volume 28, pages 802–810, 2015. (Cited on pages 15, 50 and 59.)

Shin, Woncheol; Lee, Gyubok; Lee, Jiyoung; Lyou, Eunyi; Lee, Joonseok, and Choi, Edward. Exploration into translation-equivariant image quantization. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023. (Cited on page 138.)

Shrestha, P.; Sulis, M.; Masbou, M.; Kollet, S., and Simmer, C. A scale-consistent terrestrial systems modeling platform based on cosmo, clm, and parflow. *Monthly Weather Review*, 142(9):3466 – 3483, 2014. doi: 10.1175/MWR-D-14-00029.1. URL https://journals.ametsoc.org/view/journals/mwre/142/9/mwr-d-14-00029.1.xml. (Cited on pages 21 and 79.)

Siebert, Stefan and Döll, Petra. Quantifying blue and green virtual water contents in global crop production as well as potential production losses without irrigation. *Journal of Hydrology*, 384 (3):198–217, 2010. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2009.07.031. URL https://www.sciencedirect.com/science/article/pii/S0022169409004235. Green-Blue Water Initiative (GBI). (Cited on page 216.)

Sillmann, Jana; Thorarinsdottir, Thordis; Keenlyside, Noel; Schaller, Nathalie; Alexander, Lisa V.; Hegerl, Gabriele; Seneviratne, Sonia I.; Vautard, Robert; Zhang, Xuebin, and Zwiers, Francis W. Understanding, modeling and predicting weather and climate extremes: Challenges and opportunities. *Weather and Climate Extremes*, 18:65–74, 2017. ISSN 2212-0947. doi: https://doi.org/10.1016/j.wace.2017.10.003. URL https://www.sciencedirect.com/science/article/pii/S2212094717300440. (Cited on pages 43 and 108.)

Simonyan, Karen and Zisserman, Andrew. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations (ICLR)*, 2015. (Cited on pages 14 and 87.)

Smith, Jimmy T. H.; Warrington, Andrew, and Linderman, Scott W. Simplified state space layers for sequence modeling. In *The Eleventh International Conference on Learning Representations (ICLR)*, 2023. (Cited on pages 18, 19 and 155.)

Song, Yalan; Bindas, Tadd; Shen, Chaopeng; Ji, Haoyu; Knoben, Wouter J. M.; Lonzarich, Leo; Clark, Martyn P.; Liu, Jiangtao; van Werkhoven, Katie; Lamont, Sam; Denno, Matthew; Pan, Ming; Yang, Yuan; Rapp, Jeremy; Kumar, Mukesh; Rahmani, Farshid; Thébault, Cyril; Adkins, Richard; Halgren, James; Patel, Trupesh; Patel, Arpita; Sawadekar, Kamlesh Arun, and Lawson, Kathryn. High-resolution national-scale water modeling is enhanced by multiscale differentiable physics-informed machine learning. *Water Resources Research*, 61(4):e2024WR038928, 2025. doi: https://doi.org/10.1029/2024WR038928. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024WR038928. e2024WR038928 2024WR038928. (Cited on page 48.)

Stein, Gideon; Shadaydeh, Maha; Blunk, Jan; Penzel, Niklas, and Denzler, Joachim. Causalrivers–scaling up benchmarking of causal discovery for real-world time-series. In *The Thirteenth International Conference on Learning Representations (ICLR)*, 2025. (Cited on page 47.)

Steinfeld, Daniel. *Calculation of indices for forest fire risk assessment in weather and climate data*, 2022. URL https://github.com/steidani/FireDanger. (Cited on page 20.)

Stott, Peter A.; Christidis, Nikolaos; Otto, Friederike E. L.; Sun, Ying; Vanderlinden, Jean-Paul; van Oldenborgh, Geert Jan; Vautard, Robert; von Storch, Hans; Walton, Peter; Yiou, Pascal, and Zwiers, Francis W. Attribution of extreme weather and climate-related events. *WIREs Climate Change*, 7(1):23–41, 2016. doi: https://doi.org/10.1002/wcc.380. URL https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcc.380. (Cited on page 108.)

Strothmann, Laurenz; Rascher, Uwe, and Roscher, Ribana. Detection of anomalous grapevine berries using all-convolutional autoencoders. In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 3701–3704, 2019. doi: 10.1109/IGARSS.2019.8898366. (Cited on page 44.)

Sultani, Waqas; Chen, Chen, and Shah, Mubarak. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (CVPR)*, pages 6479–6488, 2018. (Cited on pages 45, 117, 118, 128, 130 and 131.)

Sun, Alexander Y.; Li, Zhi; Lee, Wonhyun; Huang, Qixing; Scanlon, Bridget R., and Dawson, Clint. Rapid flood inundation forecast using fourier neural operator. In *IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 3735–3741, 2023. doi: 10.1109/ICCVW60793.2023.00401. (Cited on page 47.)

Sundararajan, Mukund; Taly, Ankur, and Yan, Qiqi. Axiomatic attribution for deep networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ICML'17, page 3319–3328. JMLR.org, 2017. (Cited on pages 117 and 128.)

Sushko, Vadim; Schönfeld, Edgar; Zhang, Dan; Gall, Juergen; Schiele, Bernt, and Khoreva, Anna. Oasis: Only adversarial supervision for semantic image synthesis. *International Journal of Computer Vision*, 2022. (Cited on page 48.)

Tan, Yihua; Xiong, Shengzhou, and Yan, Pei. Multi-branch convolutional neural network for built-up area extraction from remote sensing image. *Neurocomputing*, 396:358–374, 2020. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2018.09.106. URL https://www.sciencedirect.com/science/article/pii/S0925231219309208. (Cited on page 49.)

Tan, Zhentao; Chen, Dongdong; Chu, Qi; Chai, Menglei; Liao, Jing; He, Mingming; Yuan, Lu; Hua, Gang, and Yu, Nenghai. Efficient semantic image synthesis via class-adaptive normalization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):4852–4866, 2022. doi: 10.1109/TPAMI.2021.3076487. (Cited on page 48.)

Tang, Yehui; Han, Kai; Guo, Jianyuan; Xu, Chang; Li, Yanxi; Xu, Chao, and Wang, Yunhe. An image patch is a wave: Phase-aware vision mlp. In *2022 IEEE/CVF Conference on Computer Vision and*

Pattern Recognition (CVPR), pages 10925–10934, 2022a. doi: 10.1109/CVPR52688.2022.01066. (Cited on page 88.)

Tang, Yigang; Xie, Xiaolan, and Yu, Youhua. Hyperspectral classification of two-branch joint networks based on gaussian pyramid multiscale and wavelet transform. *IEEE Access*, 10:56876–56887, 2022b. doi: 10.1109/ACCESS.2022.3172501. (Cited on page 49.)

Tárraga, Jose María; Sevillano-Marco, Eva; Muñoz-Marí, Jordi; Piles, María; Sitokonstantinou, Vasileios; Ronco, Michele; Miranda, María Teresa; Cerdí, Jordi, and Camps-Valls, Gustau. Causal discovery reveals complex patterns of drought-induced displacement. *iScience*, 27(9):110628, 2024. ISSN 2589-0042. doi: https://doi.org/10.1016/j.isci.2024.110628. URL https://www.sciencedirect.com/science/article/pii/S2589004224018534. (Cited on page 3.)

Tatem, Andrew J. Worldpop, open data for spatial demography. *Scientific data*, 4(1):1–4, 2017. (Cited on page 20.)

Thakare, Kamalakar Vijay; Raghuwanshi, Yash; Dogra, Debi Prosad; Choi, Heeseung, and Kim, Ig-Jae. Dyannet: A scene dynamicity guided self-trained video anomaly detection network. In *Proceedings of the IEEE/CVF Winter conference on applications of computer vision*, pages 5541–5550, 2023. (Cited on page 44.)

Thielen, J.; Bartholmes, J.; Ramos, M.-H., and de Roo, A. The european flood alert system – part 1: Concept and development. *Hydrology and Earth System Sciences*, 13(2):125–140, 2009. doi: 10.5194/hess-13-125-2009. URL https://hess.copernicus.org/articles/13/125/2009/. (Cited on page 219.)

Thompson, Matthew P; Wei, Yu; Calkin, David E; O'Connor, Christopher D; Dunn, Christopher J; Anderson, Nathaniel M, and Hogland, John S. Risk management and analytics in wildfire response. *Current Forestry Reports*, 5(4):226–239, 2019. (Cited on page 52.)

Tian, Yu; Pang, Guansong; Chen, Yuanhong; Singh, Rajvinder; Verjans, Johan W., and Carneiro, Gustavo. Weakly-supervised video anomaly detection with robust temporal feature magnitude learning. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4955–4966, 2021. doi: 10.1109/ICCV48922.2021.00493. (Cited on pages 45, 117, 118, 128, 130 and 131.)

Tibau, Xavier-Andoni; Reimers, Christian; Gerhardus, Andreas; Denzler, Joachim; Eyring, Veronika, and Runge, Jakob. A spatiotemporal stochastic climate model for benchmarking causal discovery methods for teleconnections. *Environmental Data Science*, 1:e12, 2022. doi: 10.1017/eds.2022.11. (Cited on page 3.)

Tran, Du; Bourdev, Lubomir; Fergus, Rob; Torresani, Lorenzo, and Paluri, Manohar. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, December 2015. (Cited on page 14.)

Tucker, Compton J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2):127–150, 1979. ISSN 0034-4257. doi:

10.1016/0034-4257(79)90013-0. URL https://www.sciencedirect.com/science/article/pii/0034425779900130. (Cited on page 80.)

Tuia, Devis; Schindler, Konrad; Demir, Begüm; Zhu, Xiao Xiang; Kochupillai, Mrinalini; Džeroski, Sašo; van Rijn, Jan N.; Hoos, Holger H.; Del Frate, Fabio; Datcu, Mihai; Markl, Volker; Le Saux, Bertrand; Schneider, Rochelle, and Camps-Valls, Gustau. Artificial intelligence to advance earth observation: A review of models, recent trends, and pathways forward. *IEEE Geoscience and Remote Sensing Magazine*, pages 2–25, 2024. doi: 10.1109/MGRS.2024.3425961. (Cited on page 46.)

Tuli, Shreshth; Casale, Giuliano, and Jennings, Nicholas R. TranAD: Deep Transformer Networks for Anomaly Detection in Multivariate Time Series Data. *Proceedings of VLDB*, 15(6):1201–1214, 2022. (Cited on pages 44 and 113.)

Ulyanov, Dmitry; Vedaldi, Andrea, and Lempitsky, Victor S. Instance normalization: The missing ingredient for fast stylization. *ArXiv*, abs/1607.08022, 2016. (Cited on page 48.)

Valcke, S. The oasis3 coupler: a european climate modelling community software. *Geoscientific Model Development*, 6(2):373–388, 2013. doi: 10.5194/gmd-6-373-2013. URL https://gmd.copernicus.org/articles/6/373/2013/. (Cited on page 21.)

Van Den Oord, Aaron; Vinyals, Oriol, and others, . Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017. (Cited on pages 111, 112, 137 and 138.)

Van Loon, A. F.; Stahl, K.; Di Baldassarre, G.; Clark, J.; Rangecroft, S.; Wanders, N.; Gleeson, T.; Van Dijk, A. I. J. M.; Tallaksen, L. M.; Hannaford, J.; Uijlenhoet, R.; Teuling, A. J.; Hannah, D. M.; Sheffield, J.; Svoboda, M.; Verbeiren, B.; Wagener, T., and Van Lanen, H. A. J. Drought in a human-modified world: reframing drought definitions, understanding, and analysis approaches. *Hydrology and Earth System Sciences*, 20(9):3631–3650, 2016a. doi: 10.5194/hess-20-3631-2016. URL https://hess.copernicus.org/articles/20/3631/2016/. (Cited on page 93.)

Van Loon, Anne F; Gleeson, Tom; Clark, Julian; Van Dijk, Albert IJM; Stahl, Kerstin; Hannaford, Jamie; Di Baldassarre, Giuliano; Teuling, Adriaan J; Tallaksen, Lena M; Uijlenhoet, Remko, and others, . Drought in the anthropocene. *Nature Geoscience*, 9(2):89–91, 2016b. (Cited on page 25.)

van Noord, Nanne and Postma, Eric. A learned representation of artist-specific colourisation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, Oct 2017. (Cited on page 48.)

Van Oldenborgh, Geert Jan; van Der Wiel, Karin; Kew, Sarah; Philip, Sjoukje; Otto, Friederike; Vautard, Robert; King, Andrew; Lott, Fraser; Arrighi, Julie; Singh, Roop, and van Aalst, Maarten. Pathways and pitfalls in extreme event attribution. *Climatic Change*, 166(1-2):13, 2021. doi: 10.1007/s10584-021-03071-7. (Cited on page 79.)

Van Wagner, CE; Pickett, TL, and others, . *Equations and FORTRAN program for the Canadian forest fire weather index system*, volume 33. 1985. (Cited on pages 45 and 60.)

Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N; Kaiser, Łukasz, and Polosukhin, Illia. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. (Cited on pages 15, 16, 50, 53 and 58.)

Venkataramanan, Shashanka; Peng, Kuan-Chuan; Singh, Rajat Vikram, and Mahalanobis, Abhijit. Attention guided anomaly localization in images. In *European Conference on Computer Vision*, pages 485–503. Springer, 2020. (Cited on page 44.)

Vicente-Serrano, Sergio M.; Beguería, Santiago, and López-Moreno, Juan I. A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate*, 23(7):1696 – 1718, 2010. doi: 10.1175/2009JCLI2909.1. URL https://journals.ametsoc.org/view/journals/clim/23/7/2009jcli2909.1.xml. (Cited on page 79.)

Vicente-Serrano, Sergio M.; Peña-Angulo, Dhais; Beguería, Santiago; Domínguez-Castro, Fernando; Tomás-Burguera, Miquel; Noguera, Iván; Gimeno-Sotelo, Luis, and El Kenawy, Ahmed. Global drought trends and future projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 380(2238):20210285, 2022. doi: 10.1098/rsta.2021.0285. URL https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2021.0285. (Cited on page 79.)

Vischer, M. A.; Otero, N., and Ma, J. Spatially resolved rainfall streamflow modeling in central europe. *Hydrology and Earth System Sciences*, 29(20):5233–5250, 2025. doi: 10.5194/hess-29-5233-2025. URL https://hess.copernicus.org/articles/29/5233/2025/. (Cited on page 47.)

Vo, Tuong Quang; Kim, Seon-Ho; Nguyen, Duc Hai, and hyo Bae, Deg. Lstm-cm: a hybrid approach for natural drought prediction based on deep learning and climate models. *Stochastic Environmental Research and Risk Assessment*, 37:2035–2051, 2023. doi: 10.1007/s00477-022-02378-w. (Cited on page 46.)

Vogel, J.; Rivoire, P.; Deidda, C.; Rahimi, L.; Sauter, C. A.; Tschumi, E.; van der Wiel, K.; Zhang, T., and Zscheischler, J. Identifying meteorological drivers of extreme impacts: an application to simulated crop yields. *Earth System Dynamics*, 12(1):151–172, 2021. doi: 10.5194/esd-12-151-2021. URL https://esd.copernicus.org/articles/12/151/2021/. (Cited on page 3.)

Vreugdenhil, Mariette; Greimeister-Pfeil, Isabella; Preimesberger, Wolfgang; Camici, Stefania; Dorigo, Wouter; Enenkel, Markus; van der Schalie, Robin; Steele-Dunne, Susan, and Wagner, Wolfgang. Microwave remote sensing for agricultural drought monitoring: Recent developments and challenges. *Frontiers in Water*, 4, 2022. ISSN 2624-9375. doi: 10.3389/frwa.2022.1045451. URL https://www.frontiersin.org/articles/10.3389/frwa.2022.1045451. (Cited on page 79.)

Wada, Yoshihide; de Graaf, Inge E. M., and van Beek, Ludovicus P. H. High-resolution modeling of human and climate impacts on global water resources. *Journal of Advances in Modeling Earth Systems*, 8(2):735–763, 2016. doi: 10.1002/2015MS000618. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015MS000618. (Cited on page 216.)

Wan, Boyang; Fang, Yuming; Xia, Xue, and Mei, Jiajie. Weakly supervised video anomaly detection via center-guided discriminative learning. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, 2020. doi: 10.1109/ICME46284.2020.9102722. (Cited on pages 45, 117, 118, 128, 130 and 131.)

Wan, Zhengming; Hook, Simon, and Hulley, Glynn. Mod11a2 modis/terra land surface temperature/emissivity 8-day l3 global 1km sin grid v006, 2015. URL https://doi.org/10.5067/MODIS/MOD13A2.006. (Cited on page 20.)

Wang, Chao; Jiang, Shijie; Zheng, Yi; Han, Feng; Kumar, Rohini; Rakovec, Oldrich, and Li, Siqi. Distributed hydrological modeling with physics-encoded deep learning: A general framework and its application in the amazon. *Water Resources Research*, 60(4):e2023WR036170, 2024. doi: https://doi.org/10.1029/2023WR036170. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2023WR036170. e2023WR036170 2023WR036170. (Cited on page 48.)

Wang, Fei; Lai, Hexin; Li, Yanbin; Feng, Kai; Zhang, Zezhong; Tian, Qingqing; Zhu, Xiaomeng, and Yang, Haibo. Dynamic variation of meteorological drought and its relationships with agricultural drought across china. *Agricultural Water Management*, 261:107301, 2022a. ISSN 0378-3774. doi: https://doi.org/10.1016/j.agwat.2021.107301. URL https://www.sciencedirect.com/science/article/pii/S0378377421005783. (Cited on page 108.)

Wang, Guodong; Han, Shumin; Ding, Errui, and Huang, Di. Student-teacher feature pyramid matching for anomaly detection. *arXiv preprint arXiv:2103.04257*, 2021a. (Cited on page 44.)

Wang, Guodong; Wang, Yunhong; Qin, Jie; Zhang, Dongming; Bao, Xiuguo, and Huang, Di. Video anomaly detection by solving decoupled spatio-temporal jigsaw puzzles. In *European Conference on Computer Vision*, pages 494–511. Springer, 2022b. (Cited on page 44.)

Wang, Libo; Li, Rui; Duan, Chenxi; Zhang, Ce; Meng, Xiaoliang, and Fang, Shenghui. A novel transformer based semantic segmentation scheme for fine-resolution remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022c. doi: 10.1109/LGRS.2022.3143368. (Cited on page 88.)

Wang, Libo; Li, Rui; Zhang, Ce; Fang, Shenghui; Duan, Chenxi; Meng, Xiaoliang, and Atkinson, Peter M. Unetformer: A unet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 190:196–214, 2022d. ISSN 0924-2716. doi: https://doi.org/10.1016/j.isprsjprs.2022.06.008. URL https://www.sciencedirect.com/science/article/pii/S0924271622001654. (Cited on page 88.)

Wang, Shenzhi; Wu, Liwei; Cui, Lei, and Shen, Yujun. Glancing at the patch: Anomaly localization with global and local feature comparison. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 254–263, 2021b. doi: 10.1109/CVPR46437.2021.00032. (Cited on page 44.)

Wang, Xintao; Yu, Ke; Dong, Chao, and Change Loy, Chen. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *IEEE/CVF Conference on Computer Vision*

and Pattern Recognition (CVPR), pages 606–615, 2018. doi: 10.1109/CVPR.2018.00070. (Cited on page 48.)

Wang, Yihan; Zhang, Lujun; Erichson, N. Benjamin, and Yang, Tiantian. Investigating the streamflow simulation capability of a new mass-conserving long short-term memory (mc-lstm) model across the contiguous united states. *Journal of Hydrology*, 658:133161, 2025a. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2025.133161. URL https://www.sciencedirect.com/science/article/pii/S0022169425004998. (Cited on page 47.)

Wang, Yihan; Zhang, Lujun; Yu, Annan; Erichson, N. Benjamin, and Yang, Tiantian. A deep state space model for rainfall-runoff simulations. *Water Resources Research*, 61(12):e2025WR039888, 2025b. doi: https://doi.org/10.1029/2025WR039888. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2025WR039888. e2025WR039888 2025WR039888. (Cited on page 50.)

Wasim, Syed Talal; Khattak, Muhammad Uzair; Naseer, Muzammal; Khan, Salman; Shah, Mubarak, and Khan, Fahad Shahbaz. Video-focalnets: Spatio-temporal focal modulation for video action recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13778–13789, October 2023. (Cited on page 18.)

Wei, Jiangfeng; Liu, Xiaocong, and Zhou, Botao. Sensitivity of vegetation to climate in mid-to-high latitudes of asia and future vegetation projections. *Remote Sensing*, 15(10), 2023. ISSN 2072-4292. doi: 10.3390/rs15102648. URL https://www.mdpi.com/2072-4292/15/10/2648. (Cited on page 47.)

Weirich-Benet, Elizabeth; Pyrina, Maria; Jiménez-Esteve, Bernat; Fraenkel, Ernest; Cohen, Judah, and Domeisen, Daniela I. V. Subseasonal prediction of central european summer heatwaves with linear and random forest machine learning models. *Artificial Intelligence for the Earth Systems*, 2 (2):e220038, 2023. doi: 10.1175/AIES-D-22-0038.1. URL https://journals.ametsoc.org/view/journals/aies/2/2/AIES-D-22-0038.1.xml. (Cited on page 43.)

Wilson, Margaret F. J.; O'Connell, Brian; Brown, Colin; Guinan, Janine C., and Grehan, Anthony J. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy*, 30(1-2):3–35, 2007. doi: 10.1080/01490410701295962. URL https://doi.org/10.1080/01490410701295962. (Cited on page 22.)

Woo, Sanghyun; Park, Jongchan; Lee, Joon-Young, and Kweon, In So. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018. doi: 10.1007/978-3-030-01234-2_1. (Cited on page 86.)

Wu, Jhih-Ciang; Hsieh, He-Yen; Chen, Ding-Jie; Fuh, Chiou-Shann, and Liu, Tyng-Luh. Self-supervised sparse representation for video anomaly detection. In *European Conference on Computer Vision (ECCV) 2022*, pages 729–745, Cham, 2022. Springer Nature Switzerland. ISBN 978-3-031-19778-9. (Cited on page 45.)

Wu, Taosuo; Feng, Feng; Lin, Qian, and Bai, Hongmei. A spatio-temporal prediction of ndvi based on precipitation: an application for grazing management in the arid and semi-arid grasslands. *International Journal of Remote Sensing*, 41(6):2359–2373, 2020. doi: 10.1080/01431161.2019.

1688418. URL https://doi.org/10.1080/01431161.2019.1688418. (Cited on pages 46 and 104.)

Wu, Xiaoyang; Jiang, Li; Wang, Peng-Shuai; Liu, Zhijian; Liu, Xihui; Qiao, Yu; Ouyang, Wanli; He, Tong, and Zhao, Hengshuang. Point transformer v3: Simpler faster stronger. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4840–4851, June 2024. (Cited on page 207.)

Wu, Xin; Zhang, Gui; Yang, Zhigao; Tan, Sanqing; Yang, Yongke, and Pang, Ziheng. Machine learning for predicting forest fire occurrence in changsha: An innovative investigation into the introduction of a forest fuel factor. *Remote Sensing*, 15(17), 2023. ISSN 2072-4292. doi: 10.3390/rs15174208. URL https://www.mdpi.com/2072-4292/15/17/4208. (Cited on page 45.)

Wu, Yuxin and He, Kaiming. Group normalization. In *European Conference on Computer Vision (ECCV) 2018*, pages 3–19, Cham, 2018. Springer International Publishing. ISBN 978-3-030-01261-8. doi: 10.1007/978-3-030-01261-8_1. (Cited on pages 48 and 89.)

Wyatt, Julian; Leach, Adam; Schmon, Sebastian M., and Willcocks, Chris G. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 649–655, 2022. doi: 10.1109/CVPRW56347.2022.00080. (Cited on page 44.)

Xiao, Yicheng; Song, Lin; Huang, Shaoli; Wang, Jiangshan; Song, Siyu; Ge, Yixiao; Li, Xiu, and Shan, Ying. Mambatree: Tree topology is all you need in state space model. In *Advances in Neural Information Processing Systems*, volume 37, pages 75329–75354. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/89b89c04f55ea7c7ca989992bb6a98c0-Paper-Conference.pdf. (Cited on page 50.)

Xie, Pingping; Chen, Mingyue; Yang, Song; Yatagai, Akiyo; Hayasaka, Tadahiro; Fukushima, Yoshihiro, and Liu, Changming. A gauge-based analysis of daily precipitation over east asia. *Journal of Hydrometeorology*, 8(3):607–626, 2007. (Cited on page 35.)

Xie, Pingping; Chen, M, and Shi, W. Cpc unified gauge-based analysis of global daily precipitation. In *Preprints, 24th Conf. on Hydrology, Atlanta, GA, Amer. Meteor. Soc*, volume 2, 2010. (Cited on page 35.)

Xoplaki, E.; Ellsäßer, F.; Grieger, J.; Nissen, K. M.; Pinto, J.; Augenstein, M.; Chen, T.-C.; Feldmann, H.; Friederichs, P.; Gliksman, D.; Goulier, L.; Haustein, K.; Heinke, J.; Jach, L.; Knutzen, F.; Kollet, S.; Luterbacher, J.; Luther, N.; Mohr, S.; Mudersbach, C.; Müller, C.; Rousi, E.; Simon, F.; Suarez-Gutierrez, L.; Szemkus, S.; Vallejo-Bernal, S. M.; Vlachopoulos, O., and Wolf, F. Compound events in germany in 2018: drivers and case studies. *EGUsphere*, 2023:1–43, 2023. doi: 10.5194/egusphere-2023-1460. URL https://egusphere.copernicus.org/preprints/2023/egusphere-2023-1460/. (Cited on page 100.)

Xu, Jiehui; Wu, Haixu; Wang, Jianmin, and Long, Mingsheng. Anomaly transformer: Time series anomaly detection with association discrepancy. In *International Conference on Learning Representations (ICLR)*, 2022. URL https://openreview.net/forum?id=LzQQ89U1qm_. (Cited on pages 44 and 113.)

Xu, Yonghao; Zhang, Liangpei; Du, Bo, and Zhang, Fan. Spectral–spatial unified networks for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56 (10):5893–5909, 2018. doi: 10.1109/TGRS.2018.2827407. (Cited on page 49.)

Xu, Zhengsen; Cheng, Sibo; He, Hongjie; Wang, Lanying; Sun, Wentao; Li, Jonathan, and Xu, Lincoln Linlin. Bcwildfire: A long-term multi-factor dataset and deep learning benchmark for boreal wildfire risk prediction. *arXiv preprint arXiv:2511.17597*, 2025. (Cited on page 213.)

Yang, Jianwei; Li, Chunyuan; Zhang, Pengchuan; Dai, Xiyang; Xiao, Bin; Yuan, Lu, and Gao, Jianfeng. Focal attention for long-range interactions in vision transformers. In *Advances in Neural Information Processing Systems*, volume 34, pages 30008–30022. Curran Associates, Inc., 2021a. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/fc1a36821b02abbd2503fd949bfc9131-Paper.pdf. (Cited on pages 18 and 84.)

Yang, Jianwei; Li, Chunyuan; Dai, Xiyang, and Gao, Jianfeng. Focal modulation networks. *Advances in Neural Information Processing Systems*, 35:4203–4217, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/1b08f585b0171b74d1401a5195e986f1-Paper-Conference.pdf. (Cited on pages 17, 18, 80 and 81.)

Yang, Wenze; Guo, Wei, and Kogan, Felix. Viirs-based high resolution spectral vegetation indices for quantitative assessment of vegetation health: second version. *International Journal of Remote Sensing*, 39(21):7417–7436, 2018. doi: 10.1080/01431161.2018.1470701. URL https://doi.org/10.1080/01431161.2018.1470701. (Cited on pages 25 and 26.)

Yang, Wenze; Kogan, Felix, and Guo, Wei. An ongoing blended long-term vegetation health product for monitoring global food security. *Agronomy*, 10(12), 2020. ISSN 2073-4395. doi: 10.3390/agronomy10121936. URL https://www.mdpi.com/2073-4395/10/12/1936. (Cited on pages 23, 25, 26, 47, 80, 93, 98, 115 and 223.)

Yang, Wenze; Kogan, Felix; Guo, Wei, and Chen, Yong. A novel re-compositing approach to create continuous and consistent cross-sensor/cross-production global ndvi datasets. *International Journal of Remote Sensing*, 42(16):6023–6047, 2021b. doi: 10.1080/01431161.2021.1934597. URL https://doi.org/10.1080/01431161.2021.1934597. (Cited on pages 25, 99 and 104.)

Yang, Yuan; Feng, Dapeng; Beck, Hylke E.; Hu, Weiming; Abbas, Ather; Sengupta, Agniv; Delle Monache, Luca; Hartman, Robert; Lin, Peirong; Shen, Chaopeng, and Pan, Ming. Global daily discharge estimation based on grid long short-term memory (lstm) model and river routing. *Water Resources Research*, 61(6):e2024WR039764, 2025. doi: https://doi.org/10.1029/2024WR039764. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024WR039764. e2024WR039764 2024WR039764. (Cited on page 47.)

Yang, Zhiwei; Liu, Jing; Wu, Zhaoyang; Wu, Peng, and Liu, Xiaotao. Video event restoration based on keyframes for video anomaly detection. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (CVPR)*, pages 14592–14601, 2023. (Cited on page 44.)

Yehezkel Rohekar, Raanan; Nisimov, Shami; Gurwicz, Yaniv, and Novik, Gal. From temporal to contemporaneous iterative causal discovery in the presence of latent confounders. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 39939–39950. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/yehezkel-rohekar23a.html. (Cited on page 3.)

Yi, Jihun and Yoon, Sungroh. Patch svdd: Patch-level svdd for anomaly detection and segmentation. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, November 2020. (Cited on page 44.)

Yihdego, Yohannes; Vaheddoost, Babak, and Al-Weshah, Radwan A. Drought indices and indicators revisited. *Arabian Journal of Geosciences*, 12:1–12, 2019. (Cited on page 43.)

Yin, Ke; Shu, Lifu; Cheng, Pengle; Wang, Mingyu, and Huang, Ying. Dacsa: deformable average channel and spatial attention model for wildfire prediction and drivers. *Stochastic Environmental Research and Risk Assessment*, pages 1–15, 2025. doi: 10.1007/s00477-025-03088-9. (Cited on page 62.)

Yoon, Hyung-Jin and Voulgaris, Petros G. Multi-time predictions of wildfire grid map using remote sensing local data. *IEEE International Conference on Knowledge Graph (ICKG)*, pages 365–372, 2022. (Cited on page 45.)

You, Zhiyuan; Cui, Lei; Shen, Yujun; Yang, Kai; Lu, Xin; Zheng, Yu, and Le, Xinyi. A unified model for multi-class anomaly detection. *Advances in Neural Information Processing Systems*, 35:4571–4584, 2022. (Cited on pages 44, 117, 118, 128, 130 and 131.)

Yu, Jiahui; Li, Xin; Koh, Jing Yu; Zhang, Han; Pang, Ruoming; Qin, James; Ku, Alexander; Xu, Yuanzhong; Baldridge, Jason, and Wu, Yonghui. Vector-quantized image modeling with improved VQGAN. In *International Conference on Learning Representations (ICLR)*, 2022a. URL https://openreview.net/forum?id=pfNyExj7z2. (Cited on pages 137 and 138.)

Yu, Lijun; Lezama, José; Gundavarapu, Nitesh B.; Versari, Luca; Sohn, Kihyuk; Minnen, David; Cheng, Yong; Gupta, Agrim; Gu, Xiuye; Hauptmann, Alexander G.; Gong, Boqing; Yang, Ming-Hsuan; Essa, Irfan; Ross, David A., and Jiang, Lu. Language model beats diffusion – tokenizer is key to visual generation. In *The Twelfth International Conference on Learning Representations (ICLR)*, 2024. (Cited on pages 111, 112 and 138.)

Yu, Wentao; Li, Jing; Liu, Qinhuo; Zhao, Jing; Dong, Yadong; Wang, Cong; Lin, Shangrong; Zhu, Xinran, and Zhang, Hu. Spatial–temporal prediction of vegetation index with deep recurrent neural networks. *IEEE Geoscience and Remote Sensing Letters*, 19:1–5, 2022b. doi: 10.1109/LGRS.2021.3064814. (Cited on page 46.)

Yuan, Xing; Wang, Yumiao; Ji, Peng; Wu, Peili; Sheffield, Justin, and Otkin, Jason A. A global transition to flash droughts under climate change. *Science*, 380(6641):187–191, 2023. doi: 10.1126/science.abn6301. URL https://www.science.org/doi/abs/10.1126/science.abn6301. (Cited on page 79.)

Zaheer, Muhammad Zaigham; Lee, Jin-ha; Astrid, Marcella, and Lee, Seung-Ik. Old is gold: Redefining the adversarially learned one-class classifier training paradigm. In *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition (CVPR)*, pages 14183–14193, 2020a. (Cited on page 44.)

Zaheer, Muhammad Zaigham; Mahmood, Arif; Astrid, Marcella, and Lee, Seung-Ik. Claws: Clustering assisted weakly supervised learning with normalcy suppression for anomalous event detection. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16*, pages 358–376. Springer, 2020b. (Cited on page 45.)

Zajac, Zuzanna; Revilla-Romero, Beatriz; Salamon, Peter; Burek, Peter; Hirpa, Feyera A., and Beck, Hylke. The impact of lake and reservoir parameterization on global streamflow simulation. *Journal of Hydrology*, 548:552–568, 2017. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2017.03.022. URL https://www.sciencedirect.com/science/article/pii/S0022169417301671. (Cited on page 181.)

Zängl, Günther; Reinert, Daniel; Rípodas, Pilar, and Baldauf, Michael. The icon (icosahedral non-hydrostatic) modelling framework of dwd and mpi-m: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, 141(687):563–579, 2015. doi: 10.1002/qj.2378. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2378. (Cited on pages 218 and 219.)

Zatsarynna, Olga; Bahrami, Emad; Farha, Yazan Abu; Francesca, Gianpiero, and Gall, Juergen. Manta: Diffusion mamba for efficient and effective stochastic long-term dense action anticipation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. (Cited on page 50.)

Zavrtanik, Vitjan; Kristan, Matej, and Skočaj, Danijel. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8330–8339, 2021. (Cited on page 44.)

Zeng, Jingyu; Zhang, Rongrong; Qu, Yanping; Bento, Virgílio A.; Zhou, Tao; Lin, Yuehuan; Wu, Xiaoping; Qi, Junyu; Shui, Wei, and Wang, Qianfeng. Improving the drought monitoring capability of vhi at the global scale via ensemble indices for various vegetation types from 2001 to 2018. *Weather and Climate Extremes*, 35:100412, 2022. ISSN 2212-0947. doi: 10.1016/j.wace.2022.100412. URL https://www.sciencedirect.com/science/article/pii/S2212094722000068. (Cited on page 26.)

Zeng, Jingyu; Zhou, Tao; Qu, Yanping; Bento, Virgílio; Qi, Junyu; Xu, Yixin; Li, Ying, and Wang, Qianfeng. An improved global vegetation health index dataset in detecting vegetation drought. *Scientific Data*, 10:338, 05 2023. doi: 10.1038/s41597-023-02255-3. (Cited on pages 26 and 99.)

Zhang, Binlan; Ouyang, Chaojun; Cui, Peng; Xu, Qingsong; Wang, Dongpo; Zhang, Fei; Li, Zhong; Fan, Linfeng; Lovati, Marco; Liu, Yanling, and Zhang, Qianqian. Deep learning for cross-region streamflow and flood forecasting at a global scale. *The Innovation*, 5(3):100617, 2024a. ISSN 2666-6758. doi: https://doi.org/10.1016/j.xinn.2024.100617. URL https://www.sciencedirect.com/science/article/pii/S2666675824000559. (Cited on page 47.)

Zhang, Guoli; Wang, Ming, and Liu, Kai. Forest fire susceptibility modeling using a convolutional neural network for yunnan province of china. *International Journal of Disaster Risk Science*, 10 (3):386–403, 2019. (Cited on page 45.)

Zhang, Guoli; Wang, Ming, and Liu, Kai. Deep neural networks for global wildfire susceptibility modelling. *Ecological Indicators*, 127:107735, 2021a. ISSN 1470-160X. doi: https://doi.org/10.1016/j.ecolind.2021.107735. URL https://www.sciencedirect.com/science/article/pii/S1470160X21004003. (Cited on page 45.)

Zhang, Guowen; Fan, Lue; He, Chenhang; Lei, Zhen; Zhang, Zhaoxiang, and Zhang, Lei. Voxel mamba: Group-free state space models for point cloud based 3d object detection. In *Advances in Neural Information Processing Systems*, volume 37, pages 81489–81509. Curran Associates, Inc., 2024b. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/947b63838c90f1485188b9c673bc3a14-Paper-Conference.pdf. (Cited on page 50.)

Zhang, Jian; Kamata, Sei-ichiro, and Ueshige, Yoshifumi. A pseudo-hilbert scan algorithm for arbitrarily-sized rectangle region. In *Advances in Machine Vision, Image Processing, and Pattern Analysis*, pages 290–299, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg. ISBN 978-3-540-37598-2. doi: https://doi.org/10.1007/11821045_31. (Cited on page 159.)

Zhang, Xiang; Chen, Nengcheng; Li, Jizhen; Chen, Zhihong, and Niyogi, Dev. Multi-sensor integrated framework and index for agricultural drought monitoring. *Remote Sensing of Environment*, 188:141–163, 2017. ISSN 0034-4257. doi: https://doi.org/10.1016/j.rse.2016.10.045. URL https://www.sciencedirect.com/science/article/pii/S0034425716304242. (Cited on pages 24 and 43.)

Zhang, Ximiao; Xu, Min, and Zhou, Xiuzhuang. Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16699–16708, 2024c. (Cited on page 44.)

Zhang, Yikui; Ragettli, Silvan; Molnar, Peter; Fink, Olga, and Peleg, Nadav. Generalization of an encoder-decoder lstm model for flood prediction in ungauged catchments. *Journal of Hydrology*, 614:128577, 2022. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2022.128577. URL https://www.sciencedirect.com/science/article/pii/S0022169422011477. (Cited on page 47.)

Zhang, Yu; Hao, Zengchao; Feng, Sifang; Zhang, Xuan; Xu, Yang, and Hao, Fanghua. Agricultural drought prediction in china based on drought propagation and large-scale drivers. *Agricultural Water Management*, 255:107028, 2021b. ISSN 0378-3774. doi: https://doi.org/10.1016/

j.agwat.2021.107028. URL https://www.sciencedirect.com/science/article/pii/S0378377421002936. (Cited on page 43.)

Zhang, Zeyu; Liu, Akide; Reid, Ian; Hartley, Richard; Zhuang, Bohan, and Tang, Hao. Motion mamba: Efficient and long sequence motion generation. In *European Conference on Computer Vision (ECCV) 2024*, pages 265–282, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-73232-4. doi: https://doi.org/10.1007/978-3-031-73232-4_15. (Cited on page 50.)

Zhao, Qi; Ma, Yujing; Lyu, Shuchang, and Chen, Lijiang. Embedded self-distillation in compact multibranch ensemble network for remote sensing scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–15, 2022. doi: 10.1109/TGRS.2021.3126770. (Cited on page 49.)

Zhao, Shan; Prapas, Ioannis; Karasante, Ilektra; Xiong, Zhitong; Papoutsis, Ioannis; Camps-Valls, Gustau, and Zhu, Xiao Xiang. Causal graph neural networks for wildfire danger prediction. *arXiv preprint arXiv:2403.08414*, 2024. (Cited on page 46.)

Zheng, Yongjie; Liu, Sicong; Du, Qian; Zhao, Hui; Tong, Xiaohua, and Dalponte, Michele. A novel multitemporal deep fusion network (mdfn) for short-term multitemporal hr images classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:10691–10704, 2021. doi: 10.1109/JSTARS.2021.3119942. (Cited on page 49.)

Zheng, Yu; Koh, Huan Yee; Jin, Ming; Chi, Lianhua; Phan, Khoa T.; Pan, Shirui; Chen, Yi-Ping Phoebe, and Xiang, Wei. Correlation-aware spatial–temporal graph learning for multivariate time-series anomaly detection. *IEEE Transactions on Neural Networks and Learning Systems*, 35 (9):11802–11816, 2024. doi: 10.1109/TNNLS.2023.3325667. (Cited on page 44.)

Zhong, Jia-Xing; Li, Nannan; Kong, Weijie; Liu, Shan; Li, Thomas H., and Li, Ge. Graph convolutional label noise cleaner: Train a plug-and-play action classifier for anomaly detection. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. (Cited on page 45.)

Zhong, Liangjin; Lei, Huimin; Li, Zhiyuan, and Jiang, Shijie. Advancing streamflow prediction in data-scarce regions through vegetation-constrained distributed hybrid ecohydrological models. *Journal of Hydrology*, 645:132165, 2024. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2024.132165. URL https://www.sciencedirect.com/science/article/pii/S0022169424015610. (Cited on page 47.)

Zhong, Zhijie; Yu, Zhiwen; Yang, Yiyuan; Wang, Weizheng; Yang, Kaixiang, and Chen, C. L. Philip. Patchad: A lightweight patch-based mlp-mixer for time series anomaly detection. *IEEE Transactions on Big Data*, 11(6):3460–3473, 2025. doi: 10.1109/TBDATA.2025.3596745. (Cited on pages 44 and 113.)

Zhong, Zilong; Li, Jonathan; Ma, Lingfei; Jiang, Han, and Zhao, He. Deep residual networks for hyperspectral image classification. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 1824–1827, 2017. doi: 10.1109/IGARSS.2017.8127330. (Cited on page 49.)

Zhou, Hang; Yu, Junqing, and Yang, Wei. Dual memory units with uncertainty regulation for weakly supervised video anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 3769–3777, 2023. (Cited on page 45.)

Zhou, Yixuan; Qu, Yi; Xu, Xing; Shen, Fumin; Song, Jingkuan, and Tao Shen, Heng. Batchnorm-based weakly supervised video anomaly detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(12):13642–13654, 2024. doi: 10.1109/TCSVT.2024.3450734. (Cited on page 45.)

Zhu, Jiankai; Liu, Xiaodong; Cheng, Pengle; Wang, Mingyu, and Huang, Ying. Unveiling spatiotemporal patterns of wildfire risk: a transformer-based earth system analysis. *Climate Dynamics*, 63 (1):21, 2025. doi: 10.1007/s00382-024-07481-y. (Cited on page 46.)

Zhu, Lianghui; Liao, Bencheng; Zhang, Qian; Wang, Xinlong; Liu, Wenyu, and Wang, Xinggang. Vision mamba: Efficient visual representation learning with bidirectional state space model. In *Forty-first International Conference on Machine Learning*, 2024. (Cited on pages 50 and 158.)

Zhu, Peihao; Abdal, Rameen; Qin, Yipeng, and Wonka, Peter. Sean: Image synthesis with semantic region-adaptive normalization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5103–5112, 2020. doi: 10.1109/CVPR42600.2020.00515. (Cited on page 48.)

Zhu, Zaichun; Piao, Shilong; Myneni, Ranga B; Huang, Mengtian; Zeng, Zhenzhong; Canadell, Josep G; Ciais, Philippe; Sitch, Stephen; Friedlingstein, Pierre; Arneth, Almut; Cao, Chunxiang; Cheng, Lei; Kato, Etsushi; Koven, Charles; Li, Yue; Lian, Xu; Liu, Yongwen; Liu, Ronggao; Mao, Jiafu; Pan, Yaozhong; Peng, Shushi; Peñuelas, Josep; Poulter, Benjamin; Pugh, Thomas A. M.; Stocker, Benjamin D.; Viovy, Nicolas; Wang, Xuhui; Wang, Yingping; Xiao, Zhiqiang; Yang, Hui; Zaehle, Sönke, and Zeng, Ning. Greening of the earth and its drivers. *Nature climate change*, 6 (8):791–795, 2016. doi: 10.1038/nclimate3004. (Cited on page 79.)

Zhu, Zhuangshan; Tao, Yuxiang, and Luo, Xiaobo. Hcnnet: A hybrid convolutional neural network for spatiotemporal image fusion. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–16, 2022. doi: 10.1109/TGRS.2022.3177749. (Cited on page 49.)

Zhuang, Jiawei; raphael dussin, ; Jüling, André, and Rasp, Stephan. JiaweiZhuang/xESMF: v0.3.0 Adding ESMF.LocStream capabilities, March 2020. URL https://doi.org/10.5281/zenodo.3700105. (Cited on pages 26, 27 and 28.)

Zittis, George; Hadjinicolaou, Panos; Klangidou, Marina; Proestos, Yiannis, and Lelieveld, Jos. A multi-model, multi-scenario, and multi-domain analysis of regional climate projections for the mediterranean. *Regional Environmental Change*, 19(8):2621–2635, 2019. (Cited on page 52.)