Institute of Bio- and Geosciences 2, Forschungszentrum Jülich

A machine learning-based approach to estimate solar-induced fluorescence from airborne and spaceborne hyperspectral data

Dissertation zur

Erlangung des Grades

Doktor der Ingenieurwissenschaften (Dr.-Ing.)

der Agrar-, Ernährungs- und Ingenieurwissenschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

von

– ***Jim Loïc Buffat*** –

aus

Zürich, ZH, Schweiz

Bonn 2025

Referent**: Prof. Dr. Uwe Rascher**
Korreferent**: Prof. Dr. Cyrill Stachniss**

Tag der mündlichen Prüfung: 02.09.2025

Angefertigt mit Genehmigung der Agrar-, Ernährungs- und Ingenieurwissenschaftlichen Fakultät der Universität Bonn

Acknowledgements

I would like to thank my supervisors, **Prof. Dr. Uwe Rascher** and **Dr. Hanno Scharr**, for their guidance and support throughout this work. Their feedback and expertise have been crucial in completing this thesis.

Equally, I would like to thank my mother, **Nathalie Buffat**, and my father, **Chrigi Seiler**, whose invaluable support during so many years have allowed me to tackle this challenge.

Abstract

Techniques to estimate sun-induced fluorescence (SIF) passively from hyperspectral remote sensing data have evolved steadily over the last two decades. SIF provides causally grounded information about the photosynthetic activity of plants and, as such, is considered a valuable quantity for agriculture-related applications and various ecosystem monitoring setups. Accordingly, interest in this quantity has grown as the precision and availability of SIF estimates has improved worldwide. Upcoming missions such as FLEX by the European Space Agency (ESA) are expected to further reinforce this trend by offering, for the first time, spatially well-resolved estimates from space, based on measurements conducted in regular repeat cycles.

The growing amount of data sources that qualify for the derivation of SIF estimates, combined with an anticipated increased reliance of various stakeholders on precise Remote Sensing SIF estimates highlight the need for SIF retrieval methods with high validation accuracy applicable in a wide range of observational conditions. In response to this challenge, this thesis includes four sequential publications that develop a novel machine learning-based approach to estimate SIF in the O_2-A absorption band from airborne and spaceborne hyperspectral imagery. The proposed approach leverages recent developments in the field of deep learning for data-driven and physically consistent SIF estimation using data from HyPlant, the airborne demonstrator for FLEX, and DESIS, a spaceborne hyperspectral sensor with reduced spectral resolution.

Publication I establishes the basis for a new self-supervised neural network-based approach targeting SIF retrieval in the O_2-A absorption band of hyperspectral data from the airborne HyPlant sensor. To achieve this, a reconstruction-based loss is employed to train a multi-layer perceptron to predict the spectral decomposition of the at-sensor radiance using a physical simulation layer in the network. Despite the approximate nature of this physical model – shown to yield partially inconsistent spectral reconstructions – the method demonstrates competitive performance against in-situ top-of-canopy SIF measurements.

To address the limitations of Publication I, a closer integration of exact radiative transfer models such as MODTRAN6 in the training process targeted. The computational cost of such a model is, however, prohibitive in the training setting of artificial neural networks. Publication II therefore investigates the derivation of machine learning surrogate models that balance training and inference times with the simulation precision required for SIF retrieval in the O_2-A band. As a first application, Publication III integrates the results of this study in the general SIF retrieval framework developed in Publication I achieving state-of-the-art validation performance on HyPlant data, demonstrating strong agreement with in-situ SIF measurements.

Finally, Publication IV applies this approach to spaceborne hyperspectral data from the DESIS sensor – marking the first successful retrieval of SIF from space at 30 m resolution using a sensor previously considered unsuitable for such estimates. To validate this exceptional result, Publication IV makes use of simultaneous overflights of HyPlant and DESIS to obtain high-quality SIF estimates as reference data.

In summary, the four publications of this thesis make a significant contribution to the research field of SIF retrieval by introducing a novel and extensible framework for estimating SIF from both airborne and spaceborne hyperspectral imagery. Adaptable to upcoming data sources such as FLEX and capable of handling challenging observational conditions, including scenarios with strongly variable topography, this framework may represent a valuable addition to existing methods currently under evaluation for future FLEX data processing.

Zusammenfassung

Techniken zur Schätzung sonneninduzierter Fluoreszenz (SIF) aus passiv erfassten hyperspektralen Fernerkundungsdaten haben sich in den letzten zwei Jahrzehnten stetig weiterentwickelt. SIF liefert wertvolle Informationen über die photosynthetische Aktivität von Pflanzen und gilt daher als Schlüsselressource für agrarwissenschaftliche Anwendungen und das Ökosystem-Monitoring. Die zunehmende Genauigkeit und Verfügbarkeit von SIF-Schätzungen hat das wissenschaftliche Interesse an SIF verstärkt. Die bevorstehende FLEX-Mission der Europäischen Weltraumorganisation (ESA) wird dieses voraussichtlich weiter verstärken, da sie erstmals präzise, räumlich hochaufgelöste SIF-Schätzungen aus dem All ermöglicht.

Die zunehmende Zahl an Datenquellen, die für die Schätzung von SIF geeignet sind, sowie die erwartete verstärkte Nutzung präziser Fernerkundungsdaten durch verschiedene Interessengruppen, unterstreichen die Notwendigkeit von SIF Retrieval-Methoden mit hoher Validierungsgenauigkeit, die unter einer Vielzahl von Beobachtungsbedingungen anwendbar sind. Um diese Herausforderungen anzugehen, enthält diese Dissertation vier aufeinander aufbauende Publikationen, die einen neuartigen, auf maschinellem Lernen basierenden Ansatz zur Schätzung von SIF im O_2-A-Absorptionsband in luft- und satellitengestützter hyperspektralen Daten entwickeln. Der vorgeschlagene Ansatz nutzt aktuelle Entwicklungen im Bereich des Deep Learning für datengetriebene und physikalisch konsistente SIF-Schätzungen von Hyperspektraldaten von HyPlant – dem *airborne demonstator* von FLEX – sowie DESIS, einem satellitenbasierten Sensor mit geringerer spektraler Auflösung.

Publikation I präsentiert einen neuen Ansatz zur SIF-Schätzung im O_2-A-Band, der erstmals auf selbstüberwachten neuronalen Netzwerken basiert. Dabei wird eine rekonstruktionsbasierte Verlustfunktion dazu verwendet, eine spektrale Zerlegung der von diesen Sensoren gemessenen Strahldichte durchzuführen. Trotz approximativer physikalischer Modellierung zeigt diese Methode eine vielversprechende Validierungsleistung.

Um die Einschränkungen von Publikation I anzugehen, wird eine engere Integration exakter Strahlungstransfermodelle in den Trainingsprozess angestrebt. Da jedoch lange Simulationszeiten eine direkte Einbindung solcher Modelle in das Training neuronaler Netzwerke verunmöglichen, untersucht Publikation II verschiedene maschinelle Lernverfahren als Surrogatmodelle, die die Simulationszeit reduzieren und gleichzeitig die für SIF-Schätzungen im O_2-A-Band erforderliche Genauigkeit erhalten. Als erste Anwendung nimmt Publikation III diese Ergebnisse in die allgemeine SIF-Retrieval-Methodik aus Publikation I auf und erreicht damit eine hohe Validierungsgenauigkeit in einem HyPlant-Datensatz.

Schließlich wird der neuartige Ansatz in Publikation IV auf Daten des satllitengestützten DESIS-Sensors angewendet. Er ermöglicht erstmals eine erfolgreiche Ableitung von SIF aus satellitengestützten Daten mit einer räumlichen Auflösung von 30 m mithilfe eines Sensors, der zuvor für SIF-Schätzungen ungeeignet galt. Zur Validierung dieses erfolgreichen Ergebnisses nutzt Publikation IV koinzidente Überflüge von HyPlant und DESIS, um Referenzdaten bereitzustellen.

Die vier Publikationen dieser Dissertation leisten einen wichtigen Beitrag zur Forschung an SIF-Retrieval-Verfahren, indem sie eine neuartige Methodik zur Schätzung von SIF aus luft- und satellitengestützten hyperspektralen Bilddaten beschreiben. Da diese Methodik anpassbar an neue Sensoren wie FLEX ist und mit herausfordernden Beobachtungsbedingungen – einschließlich stark variierender Topografie – umgehen kann, stellt sie eine wertvolle Ergänzung zu bestehenden Verfahren dar, die derzeit für die künftige Verarbeitung von FLEX-Daten evaluiert werden.

Contents

List of Publications

Publications and Preprints included in this Thesis

I. **J. Buffat**, M. Pato, K. Alonso, S. Auer, E. Carmona, S. Maier, R. Müller, P. Rademske, U. Rascher, H. Scharr, B. Siegmann, "A Multi-Layer Perceptron Approach for SIF Retrieval in the O2-A Absorption Band from Hyperspectral Imagery of the Airborne HyPlant Sensor System", Remote Sensing of Environment, Volume 318, 2025, 114596, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2024.114596.

II. M. Pato, **J. Buffat**, K. Alonso, S. Auer, E. Carmona, S. Maier, R. Müller, P. Rademske, U. Rascher, H. Scharr, "Physics-based Machine Learning Emulator of at-Sensor Radiances for Solar-Induced Fluorescence Retrieval in the O_2-A Absorption Band", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2024), doi: 10.1109/JSTARS.2024.3457231

III. **J. Buffat**, M. Pato, K. Alonso, S. Auer, E. Carmona, S. Maier, R. Müller, P. Rademske, U. Rascher, H. Scharr, "Emulation-based self-supervised SIF retrieval in the O_2-A absorption band with HyPlant", submitted to Remote Sensing of Environment, preprint available, https://doi.org/10.22541/essoar.174000855.50541566/v2

IV. **J. Buffat**, M. Pato, K. Alonso, S. Auer, E. Carmona, S. Maier, R. Müller, P. Rademske, U. Rascher, H. Scharr, "Retrieval of sun-induced plant fluorescence in the O_2-A absorption band from DESIS imagery", accepted for ECCV Workshop Proceedings (2024), preprint available: https://doi.org/10.48550/arXiv.2411.08925

Other Journal Publications

V. M. Pato, K. Alonso, **J. Buffat**, S. Auer, E. Carmona, S. Maier, R. Müller, P. Rademske, U. Rascher, H. Scharr, "Simulation Framework for Solar-Induced Fluorescence Retrieval and Application to DESIS and HyPlant", submitted to Remote Sensing of Environment, preprint available, DOI: 10.2139/ssrn.5148562

VI. O. Regaieg, Z. Malenovsky, B. Siegmann, **J. Buffat**, J. Krämer, N. Lauret, V. Le Dantec, "DART-based Temporal and Spatial Retrievals of Solar-Induced Chlorophyll Fluorescence Quantum Efficiency from *in-situ* and Airborne Crop Observations", Remote Sensing of Environment, Volume 319, 2025, 114636, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2025.114636.

VII. S. Trim, **J. Buffat**, and A. Hueni, "Simulation of a simultaneous traceable spectroradiometric calibration of an imaging spectrometer," Appl. Opt. 64, 782-796 (2025), DOI: 10.1364/AO.547144

List of Conference Contributions

I. Miguel Pato, **Jim Buffat**, Kevin Alonso, Stefan Auer, Emiliano Carmona, Stefan Maier, Rupert Müller, Patrick Rademske, Uwe Rascher and Hanno Scharr, “FluoMap: Retrieving Sun-induced fluorescence from space using machine learning”, poster and lightning talk at Helmholtz AI Conference 2022, 2-3 June 2022, Dresden, Germany.

II. **Jim Buffat**, Uwe Rascher and Hanno Scharr, “A Multi-Layer Perceptron Based Regressor for SIF Retrieval from Hyperspectral Imagery of the Airborne HyPlant Sensor in Topographically Variable Terrain”, talk at EARSeL Conference 2022, 22-24 June 2022, Potsdam, Germany.

III. **Jim Buffat**, Miguel Pato, Kevin Alonso, Stefan Auer, Emiliano Carmona, Stefan Maier, Rupert Müller, Patrick Rademske, Uwe Rascher and Hanno Scharr, “Emulator-based neural network prediction for SIF retrieval in the O2-A absorption band”, presentation at Helmholtz AI conference, 12-14 Jun 2023, Hamburg, Germany.

IV. **Jim Buffat**, Miguel Pato, Kevin Alonso, Emiliano Carmona, Stefan Maier, Rupert Müller, Stefan Auer, Patrick Rademske, Uwe Rascher and Hanno Scharr, “Simulation and Emulation of HyPlant and DESIS Spectra in the O2-A Absorption Band for a novel Neural Network based Self-Supervised SIF Retrieval Approach”, poster at FLEX Workshop, 19-21 Sep 2023, Frascati, Italy.

V. **J. Buffat**, M. Pato, K. Alonso, S. Auer, E. Carmona, S. Maier, R. Müller, P. Rademske, U. Rascher, H. Scharr, "Deep Learning Based Prediction of Sun-Induced Fluorescence from Hyplant Imagery," IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 2023, pp. 2993-2996, doi: 10.1109/IGARSS52108.2023.10282828.

VI. M. Pato, K. Alonso, S. Auer, **J. Buffat**, E. Carmona, S. Maier, R. Müller, P. Rademske, U. Rascher, H. Scharr, "Fast Machine Learning Simulator of At-Sensor Radiances for Solar-Induced Fluorescence Retrieval with DESIS and Hyplant," IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 2023, pp. 7563-7566, doi: 10.1109/IGARSS52108.2023.10281579.

VII. **Jim Buffat**, Miguel Pato, Kevin Alonso, Stefan Auer, Emiliano Carmona, Stefan Maier, Rupert Müller, Patrick Rademske, Uwe Rascher and Hanno Scharr, “Leveraging a large-scale radiative transfer simulation for an emulator based retrieval scheme of sun- induced fluorescence in HyPlant imagery”, talk at EARSeL Conference 2022, 16-18 June 2024, Valencia, Spain.

VIII. Miguel Pato, **Jim Buffat**, Kevin Alonso, Stefan Auer, Emiliano Carmona, Stefan Maier, Rupert Müller, Patrick Rademske, Uwe Rascher and Hanno Scharr, “Towards A Machine Learning Retrieval Of Solar-induced Fluorescence From DESIS Data”, talk at EARSeL Conference 2024 16-18 June 2024, Valencia, Spain.

IX. **Jim Buffat**, "Coupling a radiative transfer emulator and self-supervised neural network training for the retrieval of sun-induced fluorescence from hyperspectral imagery", talk in ELLIIT Focus Seminar Series as a visiting Scholar, 15 October 2024, Linköping, Sweden

X. **Jim Buffat**, Miguel Pato, Kevin Alonso, Stefan Auer, Emiliano Carmona, Stefan Maier, Rupert Müller, Patrick Rademske, Uwe Rascher and Hanno Scharr, "Towards fast and sensor-independent retrieval of sun-induced fluorescence from spaceborne hyperspectral data", poster at 3rd Workshop on International Cooperation in Spaceborne Imaging Spectroscopy 2024, 13-15 November 2024, ESTEC, Noordwijk, Netherlands

Chapter 1

Introduction

1.1 Overview

The recent decades of research in optical remote sensing applications could rely on a steadily increased number of airborne and spaceborne hyperspectral sensors with improved spectral coverage, spatial and spectral resolution and signal-to-noise ratio [231]. The high number of spectral features recorded in parallel over large spatial footprints by optical remote sensing have become crucial information sources in a wide array of applications related to global environmental mapping and monitoring [64, 74]. The amount of information gathered from airborne and satellite-based optical remote sensing sensors has increased significantly over the last decades [101] and will likely grow further over the next decade to fulfill the expanding interest in remote sensing derived estimates of terrestrial variables [63, 186, 187] which are crucial for climate change monitoring, adaptation and mitigation strategies [160, 259, 266]. The usefulness of hyperspectral imagery for environmental and agricultural monitoring in particular has been recognized early in the development of hyperspectral remote sensing systems [105, 274] and is a significant driver in the design and application of new spaceborne hyperspectral sensors [13, 17, 40, 210, 280]. Recently, the success of deep learning (DL) in standard imagery, as well its application in optical multispectral remote sensing a number of publications have shown the increased potential for hyperspectral imagery for providing accurate estimates of surface quantities [103, 118] to enable proactive, data-driven strategies to anticipate, monitor, and respond to climate change-induced societal challenges worldwide.

This thesis focuses on the use of deep learning techniques for the estimation of sun-induced fluorescence (SIF) from hyperspectral imagery. The estimation of SIF has accompanied developments in hyperspectral remote sensing and imaging since its beginnings [194]. To date, SIF estimation still poses several technical and methodological issues even though today it can rely on powerful computational hardware and a growing number of sensors with adapted instrument characteristics. The research summarized in this thesis targets some of these challenges. The publications included in this thesis were conducted in the framework of the Helmholtz AI project *FluoMap* (2021 - 2024) where for the first time a purely DL-based approach for SIF estimation was targeted. To this end, *FluoMap* has focused on two representative sensors – HyPlant FLUO and DESIS. The Publications I - IV [29, 31, 33, 213] included in this thesis focus on methodological aspects of SIF estimation. In addition, further publications focusing on hyperspectral signal simulation (Publication V [212]) and emulation [211] as well as on SIF estimation [30] have been prepared for *FluoMap*.

In the following Chapter 1, this thesis presents a concise overview of the current state of research on SIF estimation. Subsequently, it examines the contributions of Publications I - IV to the field of hyperspectral SIF estimation. To this end, summaries

of Publications I - IV are provided in Chapter 2. A synthesizing Discussion of these four Publications is given in Chapter 3. The thesis concludes with giving a summary over future challenges and perspectives for further research in deep learning based SIF retrieval both in airborne and spaceborne contexts in Chapter 4. For the reader's convenience, publications authored or coauthored by the author of this thesis are referred to by their publication numbers (I - VII), as listed in the *List of Publications*, in addition to their full bibliographic references.

1.2 Background and Motivation

1.2.1 Chlorophyll Fluorescence

Chlorophyll *a* fluorescence (ChlF) is a light source in the red and far-red spectrum (600 - 770 nm) that originates in the chloroplast of plants, algae and cyanobacteria [207]. It is one of three de-excitation pathways by which the solar irradiance absorbed by vegetative tissues is consumed and channeled to prevent photoinduced damage to the plant. The energy absorbed by plants is either quenched by photochemical processes (PQ) fuelling the linear electron transport of the photosynthetic apparatus, by non-phochemical processes (NPQ) that converts it to thermal energy or else by re-emission in the form of ChlF [188]. Under normal illumination conditions, healthy plants are able to regulate the amount of energy dissipated in the photochemical and non-photochemical pathway [291] by adjusting their metabolism and physiological state (e.g., by opening and closing of stomata) [196]. ChlF itself is, thus, not directly regulated by plants [278] but represents the residual energy that is not otherwise dissipated via photochemical or non-photochemical quenching.

1.2.2 Active Chlorophyll Fluorescence Measurements

Typically, the excess energy dissipated in form of ChlF represents 2-5% of the absorbed photosynthetically active radiation (PAR). Under normal solar illumination conditions of vegetation, this represents a radiation intensity that can be measured passively with modern high precision spectrometer equipment. Until the year 2000, ChlF measurements were, however, measured almost exclusively actively due to lacking convincing evidence that the sun-induced fluorescence (SIF) signal could be disentangled with high-precision from the much larger background radiance [194]. A range of techniques have been developed to derive a plant's photosynthetic state from precise measurements of ChlF when it is actively induced by LEDs or lasers in field conditions [335]. Pulse-amplitude modulation (PAM) devices [100, 228] and Laser-induced Fluorescence Transient (LIFT) setups [12, 219] in particular have proven to be active measurement techniques adapted to field conditions. In both cases, a measurement series of ChlF emissions induced by repeated modulated or transient excitations is acquired. A set of variables can be derived from these time series measurements that are directly related to variables of photosynthetic activity in theoretical models of photosynthesis. The continuous deployment of these active setups allows to track the photosynthetic response of plants to changing conditions at short temporal intervals. Due to the sensitivity of the photosynthetic apparatus to biotic and abiotic stress factors causing suboptimal plant functioning and, consequently, changing photosynthetic efficiency, the continuous deployment of these measurement techniques has been historically and still is used widely in crop and yield improvement studies for breeding [94].

1.2.3 Passive Chlorophyll Fluorescence Estimation with Remote Sensing Techniques

Currently, active ChlF inducing methods cannot be applied on spatial scales larger than a few square meters except for very specific cases where the measurement deployment is highly automated [150]. ChlF emissions can be estimated from passively acquired radiance data over larger areas as sun-induced fluorescence (SIF) with suitable hyperspectral sensors. Instead of relying on controlled light excitation and fixed measurement protocols as in active measurement setups, passive measurement techniques aim at quantifying steady-state SIF that is emitted "*[...] from a photosynthetically active plant adapted to ambient irradiance when the electron transport processes and the coupled biochemical reactions of the carbon reduction cycle establish a dynamic equilibrium*" [183]. The connection between steady-state ChlF and plant physiology is subject to active research. Several works have shown that a link can be established using mechanistic photosynthesis models [109, 229, 278, 291] relying on strong assumptions and complementing the analysis with ancillary information that often may be gained from the same remote sensing data sources on which the SIF estimation is based.

Passive methods estimate the fluorescence contribution to radiance measurements in a process that is called *SIF retrieval*. While passive SIF estimates provide the closest causal link to photosynthetic activity that can be remotely sensed, such singular, non-controlled estimates of steady-state ChlF provide a reduced degree of information about the photosynthetic state of vegetation as compared to active methods due to the following factors:

1. **Energy absorption and partitioning**: ChlF is an informative quantity if it can be related to the fraction of photosynthetically active radiation energy that drives photosynthesis (PQ). In most remote sensing setups this is, however, problematic due to lacking knowledge about (i) the total amount of photosynthetically active radiation absorbed by photosynthetic pigments (APAR) and (ii) the fraction of APAR dissipated via NPQ. APAR can be controlled in active ChlF measurements, but in typical passive remote sensing measurement setups it can only be estimated. Such estimations are challenging and typically associated with large uncertainties [47, 227]. Furthermore, since ChlF is a residual energy flux, it can only be related to PQ if NPQ is known, but while in active measurement setups, the partitioning into ChlF, PQ and NPQ can be solved exactly [195] in passive ChlF sensing it must also be estimated. Accurate estimation of NPQ is subject to active research. Multiple avenues based on hyperspectral data processing have been proposed, notably some relying on the photochemical reflectance index (PRI) [3]. The uncertainties associated with NPQ estimates are, however, still significant and impose severe limits on the direct leveraging of ChlF estimates for the characterization of photosynthetic efficiency.

2. **Canopy structure and reabsorption**: Canopy structure is one of the main drivers of remotely sensed SIF [69]. It affects (i) the penetration of the canopy by solar irradiance, (ii) the spectral distribution of canopy leaving SIF due to reabsorption and scattering of ChlF along the upwelling optical path through the canopy [227] and (iii) the relative contribution of diffuse and direct irradiation [20]. Remote sensing SIF estimates therefore only relate to the top-of-canopy

(TOC) leaving SIF emission. These estimates must be post-processed with additional information describing the canopy to gain physiologically relevant leaf-level SIF [242, 264, 323]. An additional complication is given by the non-isotropic nature of differential penetration, scattering and reabsorption that leads to a viewing angle dependency in estimated SIF [167, 294, 348, 349] for which a number of normalizing schemes have been developed [122, 175, 349] (see Section 1.5.3).

3. **Atmospheric absorption**: depending on the retrieval algorithm, the spectral range of at-sensor radiance from which SIF is estimated is modulated by scattering and absorption of atmospheric components such as the oxygen and water vapour column and aerosol concentration along the optical path (see Publication II [213], Publication V [212]). These modulating effects must be accounted for under typically incomplete knowledge about individual atmospheric components. This process – called *atmospheric correction* if it is conducted as a preprocessing step and *atmospheric characterization* if performed as part of the retrieval – is a major source of the total TOC SIF estimation uncertainty (see Section 1.5.2).

4. **Pixel-level mixing**: there is a representation mismatch between pixel-level estimation and plant-level photosynthesis due to the finite resolution of remote sensing sensors. The correlation between leaf-level and TOC SIF decreases as a function sub-pixel heterogeneity in reflectance, canopy structure and soil conditions. This has been discussed for airborne [35, 128, 334] and spaceborne [98, 133, 246] retrieval contexts. Due to variable canopy reabsorption in different canopy structures, SIF mixing in single pixels can be non-linear especially in the red domain where the SIF canopy interaction is stronger than in the NIR spectral range [247].

1.2.4 SIF Applications

A range of remote sensing platforms suitable for SIF retrieval have become available in the last decades [194]. These platforms cover spatial scales from the field to the global level. The rapid growth of the hyperspectral remote sensing domain will likely allow for increased passive SIF retrieval capabilities from Unmanned Aerial Vehicles (UAVs), airplanes and spaceborne platforms in the future (see Sec. section 1.3). The scientific and economic domains benefiting from passively sensed SIF vary strongly with the spatial resolution and the uncertainty of the estimates that may be gained from the acquired remote sensing data.

A large range of remote sensing applications has been put forward that make use of SIF either based on mechanistic models of the ChlF generating processes or purely statistical approaches. Especially in statistical approaches the basis for including SIF in various inference tasks is its close causal link to the photosynthetic state of vegetation that is very sensitive to changes in environmental conditions. In the following a summary over the most researched applications of SIF estimates [279] from airborne and spaceborne platforms is given.

Gross Primary Productivity

Current methodologies for estimating global Gross Primary Productivity (GPP) are subject to substantial uncertainties, as evidenced by the considerable variability in

results across different estimation approaches [249] and low performance of predictions by Dynamic Global Vegetation Models (DGVM) with respect to observational data [92]. The reason for this large variance relies in a large number of unconstrained model assumptions [12] that have to be estimated due to a lack of direct measurement techniques at a global level. The sparsity of constraining measurements in traditional observation-based GPP estimates such as the light-use efficiency approach [248] and approaches to upscale localized Eddy Covariance (EC) GPP measurements [145] to a global distribution must be countered by a large number of model assumptions and ancillary environmental and climatic information.

Frankenberg et al. [90] has first compared global SIF measurements to process-based GPP estimates as an alternative and found comparable prediction performance to reflectance-based estimates without using additional information sources (e.g., climatic variables). SIF is a promising GPP predictor because ChlF and GPP are linked at the cellular level by the regulation of the light reactions of photoynthesis [200, 278]. This mechanistic link provides the basis for an alternative way to estimate GPP that relies less on modelling assumptions [109] and has been an actively researched topic over the last decade [109, 278].

Direct use of mechanistic models in the estimation of GPP from SIF has been investigated recently with different global and field-level SIF datasets [50, 120, 173]. The most widely used SIF-based GPP estimates rely on the empirical extraction of linear relations of SIF and GPP [117]. Empirically, the relationship between SIF and GPP is, however, variable across environmental conditions [59, 214, 299] resulting in global level variations as a function of biome and climate conditions [162, 260]. Combining SIF observations and dynamic vegetation modeling, Parazoo et al. [208] has cast the estimation of this variable SIF-GPP scaling in a Bayesian framework. Research in the scaling variance has included analysis of GPP-SIF scaling factors across different biomes and climate zones [163, 217, 233], under improved differentiation of NPQ conditions [311], under accounting for local relative distributions of C3 and C4 plants [111, 123] as well as a function of physiological plant state [19]. Other studies have concentrated on the assessment of the SIF-GPP scaling in specific vegetation types and ecosystems such as in maize [304], deciduous [327], boreal evergreen [180, 218, 305] and mangrove forests [355] as well as in peatlands [14].

A precise mechanistic understanding of coupling and decoupling mechanisms of the SIF-GPP relationship [226] is necessary to disentangle structural from physiological drivers of SIF and, thus, interpret the SIF-GPP relationship correctly. Field-level experiments [69, 107, 317, 324] have for example shown that the empirical SIF-GPP correlation is mainly driven by absorbed photosynthetically active PAR (APAR) and the canopy structure. Statistical approaches that estimate GPP from linear SIF scalings cannot identify such individual drivers. Moreover, there is evidence that under heat stress such linear scalings become [315] uninformative.

Another technique to leverage SIF estimates for the prediction of photosynthetic rates involves terrestrial biosphere models (TBMs) and land-surface models (LSMs) such as ORCHIDEE, BETHY-SCOPE and SCOPE [279]. The GPP estimation is achieved by optimizing TBM/LSM parameters that affect both SIF and GPP on SIF time series observations from spaceborne sensors [18, 200, 343]. While current GPP products mainly make use of scalar SIF estimates at particular wavelengths, accurate modelling of the whole SIF emission spectrum from 400 - 900 nm by SCOPE allows for retrieval formulations involving spectrally explicit radiance measurements [204, 302].

Precision Agriculture

The benefits of leveraging SIF as a predictor variable for agricultural applications has been recognized early in the history of SIF retrieval [65, 182]. The close link between ChlF emission and plant photosynthesis function forms the basis for research in SIF applications in phenotyping, crop monitoring and large scale yield prediction.

SIF retrieval from proximal hyperspectral sensors has been shown to have large potential in high-throughput phenotyping applications [44, 94, 269] over large areas that can only be captured by remote sensing instrumentation. The potential for fast screening of crop variants is of particular interest in breeding applications [221, 243, 245]. SIF estimates allow to estimate the photosynthetic capacity [39, 93] and the nitrogen content [38, 137, 286, 312] across different experimental field setups [3, 22, 106, 316]. Since SIF is a dynamic quantity that reacts instantaneously to changes in the photosynthetic state of vegetation, it allows furthermore to capture the response on a very short time scale. This is in contrast to analyses based on vegetation indices [104] calculated from reflectance measurements which ultimately reflect slowly changing pigment concentrations and structural vegetation properties in response to changing ground conditions [44, 56, 176].

The sensitivity of SIF to plant stress and photosynthetic activity has lead to various studies on the feasibility of field level crop monitoring from UAV [23, 48, 295, 307–309] and airborne [156, 332, 333] sensing platforms and regional crop yield prediction [149, 171, 216, 268]. Parallely, new ground-based point spectrometer systems for long-term SIF measurement series have been developed in the last decade such as the *Fluorescence Box* (FloX, JB Hyperspectral Devices AG, Dusseldorf, Germany) [134] and the *PhotoSpec* device [108].

The operational use of SIF in precise agriculture applications, however, is not yet feasible for generalized setups. The main difficulties to achieve this consist in linking top-of-canopy SIF to ChlF in different crop types and across non-experimental setups where constraining variables are not controlled [264]. On a larger spatial scale the benefits of SIF-based yield prediction from current spaceborne SIF products over traditional reflectance-based approaches is not well established [97, 216, 267].

Stress Detection and Monitoring

Remote monitoring of agricultural landscapes on a local, regional or global scale with optical remote sensing methods is considered an important tool in guaranteeing food security in the context of rising world population [24]. As a consequence, besides making use of the valuable information source of SIF in field-level precision farming applications, research in SIF applications focuses on the early detection of plant stress [2, 221] and tracking of global phenology [341] from individual or time-series observations.

In a large meta-analysis [2] evaluates the potential of active and passive ChlF and SIF in different spectral regions to reliably indicate nitrogen deficiency, temperature stress and water deficiency in plants. With increasing availability of SIF products, hyperspectral instruments for SIF retrieval and the improvement of SIF retrieval algorithms for data from different platforms, further studies have investigated heat [148], water [56, 67, 125, 172, 307], low-temperature [76] and biotic [36, 37, 77, 225, 237] stress in small-scale experimental setups with proximal, airborne and spaceborne sensors. SIF from spaceborne instruments have allowed the tracking of spatio-temporal patterns of drought and heatwaves [127, 234, 277, 330] on a regional to global scale.

As in other SIF applications, a key difficulty in stress detection with SIF remains the disentangling of physiological and structural changes [77, 148, 310]. Recently, several avenues to reduce the effects of confounding factors on the inversion have been proposed. [325, 339] use the full radiance and reflectance spectrum to infer structural properties and constrain the physiologically relevant contribution to measured SIF. Furthermore, Publication VI [241] has proposed a 3D Radiative Transfer modelling approach to downscale the canopy level SIF estimates gained from a traditional retrieval algorithm as well as with the retrieval approach developed in Publication I [33].

Constraining the Hydrological Cycle

There is a strong link between plant transpiration and the photosynthetic activity that leads to the emission of SIF [61, 273]. This circumstance likely explains the strong correlation of evapotranspiration and SIF that has been observed in drought studies [277]. SIF estimates have recently been proposed to be used to estimate [179, 322, 354] and constrain [205] global and field-scale [6, 60, 67] plant transpiration. Data assimilation of spaceborne SIF observations in land surface models for simultaneous GPP and transpiration estimation has been investigated by [232].

1.3 SIF Retrieval Methods

A common conceptual framework underlying all SIF retrieval methods describes a measured at-sensor radiance spectrum L in wavelengths λ as the sum of three main groups of radiative fluxes

$$L(\lambda) = \left(L_p + \frac{E_g^0 \, \rho \, T^{\uparrow}}{\pi \, (1 - \rho S)} + T^{\uparrow} L_f \right) (\lambda), \tag{1.1}$$

where the first term denotes the path radiance, the second reflected solar irradiance and the third the contribution of top-of-canopy SIF L_f to the at-sensor radiance L (see Publication II [213] and Publication V [212]). E_g^0 denotes the solar irradiance at the ground, ρ the surface reflectance, $T^{\uparrow}$ the total transmission coefficient from surface to sensor including direct and diffuse contributions and S the spherical albedo. This formulation hides the complex interaction of SIF with the canopy structure (see Section 1.2.3) and describes the target variable $L_f(\lambda)$ as the canopy leaving SIF integrated over the footprint of the spectral radiance measurement. Suitable parameterizations of the spectral shape of $L_f(\lambda)$ and its variability under the impact of plant stress [181] and canopy structure [242] from first principles [292], measurements [10] or retrieval performance assessments [54] are subject to active research. In the publications included in this thesis, L_f was modelled with a Gaussian which closely matches more exact formulations in the small small spectral window around the O_2-A absorption band (Publication V [212]). SIF retrieval methods attempt to disentangle the SIF contribution L_f from the impact of the atmospheric effects captured in $T^{\uparrow}$ and S, the path radiance L_p and the reflected solar irradiance ρE_g^0 in radiance measurements. To this end, these methodsfs ht leverage at-sensor radiance measurements in spectral regions with strong sensitivity to ChlF. These regions typically consist of solar and telluric absorption lines [194].

A number of reviews [21, 46, 88, 193, 194] have summarized the progress of different variants of SIF retrieval methods. An overview over the most widespread SIF retrieval algorithm is given in the following paragraphs.

1.3.1 Solar Fraunhofer Lines

The earliest form of SIF retrieval algorithms were designed to estimate SIF from spectrometer radiance measurements with sufficient spectral resolution (SR) to resolve solar Fraunhofer lines (FL) ($\lesssim$ 0.1 nm). In particular, the FLs Hα centered at 656.5 [166, 265], Fe I at 758.8 nm [89, 116], K I close to 770 [89, 143] and Ca II near 866 nm [142] have been considered for SIF retrieval due to their coverage of the ChlF spectrum. Atmospheric gases interact with light selectively such that at sufficient SR spectral features, Fraunhofer Lines, can be identified that are almost unaffected by the downwelling and upwelling radiative transfer through the atmosphere [328]. As a consequence, the relative fractional depth of FL can be assumed to be almost uniquely driven by the variation in ChlF which allows the use of FL for SIF retrieval. Depending on specific FL and the sensor resolution, additional factors affect the fractional depth of FL besides ChlF to a small extent. For example, the impact of Rotational Raman scattering (RRS) on FL-based SIF retrieval, which can be neglected only when focusing on FL which are sufficiently isolated from telluric O_2-A absorption lines in nadir-looking acquisitions [89, 113], has been studied extensively [142, 143, 265]. At non-optimal spectral resolutions, the fractional depth of measured at-sensor radiance in solar FL may furthermore be affected by spectrally close telluric absorption features due to mixing in the spectral integration described by the Instrument Spectral Response Function (ISRF). This is notably the case for Hα-based retrievals [168].

1.3.2 O_2-A and O_2-B Absorption Bands

The use of Fraunhofer Lines (FL) for fluorescence retrieval requires high spectral resolution (SR) and signal-to-noise ratio (SNR) of a sensor. To achieve the high SNR in particular, either long integration times or a large field of view per pixel is necessary to assure a sufficient light flux to the sensor. Both conditions limit fluorescence retrieval in conditions where high spatio-temporal resolution is of interest such as in FLEX mission [75] advanced by the European Space Agency (ESA). SIF retrieval from spectral regions with relatively high sensitivity to ChlF emission at smaller spectral resolutions are therefore key for (i) the design of spatial high-resolution SIF retrieval sensors such as FLEX and (ii) for commercially available sub-nanometer spectrometers [192]. Besides the solar absorption lines, the sensitivity of at-sensor radiance to SIF has maxima in the telluric oxygen absorption bands O_2-A and O_2-B [87, 89, 113]. These spectral features have the benefit of being broad enough to put less strict requirements on the sensors SR, SNR and spectral sampling interval (SSI) than for Fraunhofer absorption lines [57]. Since the O_2-A and O_2-B bands originate atmospheric oxygen absorption, these bands, however, vary as a function of the optical path length through the atmosphere which mainly depends on the target-observer-geometry. To a lesser extent water vapour and aerosol density affect the at-sensor radiance in these spectral windows as well. As a consequence, the SIF prediction must account for a larger variability in the radiance observation than is the case in FL-based approaches. In order to model such influences, multiple studies focusing on the variability of oxygen absorption bands have been conducted. For

example, Publication V [212] has presents a thorough sensitivity analysis of the O_2-A band under fluorescence emission and atmospheric effects in HyPlant and DESIS data as a basis to Publication II [213] included in this thesis. Smaller studies including also sensitivity studies of the O_2-B absorption band can be found in [87, 113, 252].

1.3.3 Fraunhofer Line Discrimination

The Fraunhofer Line Discrimination (FLD) method derives a simple equation system from Eq. 1.1 that can be solved with measurements or estimates of E_s^0 and L inside and outside of an atmospheric feature [165, 223]. While the original form of FLD targeted solar FL, most modern applications of FLD make use of the telluric O_2-A absorption feature. This is due to the strong requirements on SR, SNR and SSI to resolve individual solar FL which often can not be satisfied for spectrometers in operational conditions. Operational spectrometers capable of resolving FL for high-quality FLD SIF retrieval are restricted to satellite missions targeting the quantification of global concentrations of Greenhouse gases (GHG) (see Section 1.4.3). SIF retrieval on datasets of these sensors is, however, usually performed with optimization techniques (see Sec. 1.3.4).

FLD is mainly applied in updated forms such as the 3FLD [182] and the Improved FLD (iFLD) [9] in the telluric O_2-A absorption band. These updated forms (i) account for the typical spectral shape of at-sensor radiance, reflectance and top-of-canopy SIF in the O2-A and O_2-B spectral region and (ii) adapt the number of sampling points used in the equations [45, 170] to reduce retrieval errors. In order to account for the observational and atmospheric variability affecting the O_2-A and O_2-B band depth, these FLD variants either measure the atmospheric transfer or formulate the equations system in the framework of a radiative two-stream model where the atmospheric transfer is entirely parameterized by additional atmospheric estimates.

FLD-based approaches in O_2 absorption bands are mainly used for data from field spectrometers as well as for UAV and airborne sensor data but have also been considered for use for spaceborne data in the past [112, 113]. The main limitations of FLD methods are their high-sensitivity to noise [168] and simplistic modelling of reflectance and fluorescence shapes which does not allow the retrieval of the full SIF spectrum.

1.3.4 Fraunhofer Line Optimization Retrievals

SIF retrieval with spectral data from spaceborne GHG sensors is most prevalently performed with optimization methods targeting FL. These methods make use of the invariance of the fractional depth of solar FL by constraining the data to small spectral windows which allows to disregard atmospheric absorption. The retrieval can then be tackled by optimizing a simplified forward model that depends only on a high resolution model of solar top-of-atmosphere irradiance, the targeted fluorescence, a polynomial low-frequency contribution to account for variable reflectance and atmospheric scattering and a precise sensor characterization. Importantly, since the atmospheric influence in the focused spectral range is minimal, this model does not need to account explicitly for any atmospheric components. The first SIF retrieval methods of this type were applied to data of the TANSO-FTS sensor [89, 142, 143, 275]. More recently optimization has been applied to OCO-2/3 data as well [152]. Furthermore, [147] formulates a similar procedure in the framework of Differential Optical Absorption Spectroscopy (DOAS).

1.3.5 Data-driven Decomposition-based Methods

The simple physical modelling of the signal generation in the approaches above hinges on small spectral ranges. The restriction to small spectral ranges is, however, a bottleneck in sensors with lower spectral resolution and subjects the estimated SIF to more sensor noise. As a response to this, multiple authors have proposed using larger spectral windows which include strong atmospheric absorption features from O_2 and water vapour both in spaceborne [79, 114, 138, 153, 253, 293, 352], airborne [263] and field spectroscopy data [116, 342]. In order to simplify the description of the atmospheric influences on the observational data driven decomposition-based methods (DDDM) model the radiative transfer with linear combinations of Principal Components (PCs) from spectral datasets of non-fluorescing targets. The accuracy of such methods is very sensitive to a representative choice of PCs given the acquisitions from which SIF is to be estimated [46, 169], but have been shown to yield competitive validation accuracies in well-defined conditions with high spectral resolutions [197, 345].

1.3.6 Spectral Fitting Methods

Differently to DDDMs, Spectral Fitting Methods (SFMs) model accurately the physical generation process of either measured the at-sensor radiance [52, 54, 298] or the computed reflectance [42, 189, 192]. The estimation then is conducted similarly to DDDM by optimizing the free parameters in a least-squares minimization. The best estimate for the SIF parameterization is used as SIF estimate. If reflectance is the target, solar irradiance measurements are needed in addition to the at-sensor radiance. The downwelling and upwelling radiative transfer for SFM targeting radiance is mostly modelled with physical radiative transfer models such as MODTRAN [25]. The spectral shapes of reflectance and fluorescence emission are either derived from land-surface and biologically accurate models [52, 54, 292, 351] or these models are used directly during inference [42, 298].

Both reflectance and radiance targeting retrievals require atmospheric characterization or correction prior to the retrieval of surface parameters. While radiance targeting SFM include an explicit formulation of the radiative transfer, the atmospheric signal component is treated differently to the rest of the optimization. Instead of solving a joint optimization of atmospheric and surface related parameters, a two-step approach is usually adopted where the atmospheric contributions to the signal are first estimated with an RTM inversion scheme and subsequent RTM modelling called *interrogation technique* [296, 297]. In a second step, all surface related parameters are usually jointly estimated. This procedure is adopted because of the high computational burden of high-resolution RTMs that renders repeated runs in optimization iterations infeasible. Recently, the Publications I, III and IV, included in this thesis, and Buffat et al. [30] have, however, argued for a neural network-based SFM formulation where a joint estimation is made possible by the adoption of constraint-based neural network training.

1.4 Existing SIF Products

SIF is estimated from hyperspectral sensors installed on a large variety of platforms including ground-based, UAV, airborne, and spaceborne systems, each offering unique advantages and challenges that may be suited for particular downstream applications. Airborne and spaceborne SIF retrievals have been extensively reviewed,

highlighting the scientific interest in reliable and systematically provided SIF products [193, 194, 279]. UAV-based SIF studies have gained attention for their ability to bridge the gap between ground, airborne and potentially spaceborne observations, providing detailed insights into canopy-level fluorescence [21, 308]. Harmonizing data across these platforms remains a critical challenge to ensure consistency and comparability in SIF estimates across scales. The following section introduces the characteristics, advancements, and integration efforts regarding existing SIF products across different platforms.

1.4.1 Field and Tower-based Measurements

A number of commercially available hyperspectral spectroradiometers have been found suitable [144] to be integrated in automated measurement systems for quantitative measurements of TOC SIF in field conditions either mounted on portable stands, existing Eddy Covariance (EC) towers or UAVs [43, 80]. In the last decade, a number of research institutions have developed automated TOC SIF measurement systems (HSI [191], FluoSpec [327], FluoSpec2 [326], MRI [53], PhotoSpec [108], FAME [110], FLoX [134]). These automated systems provide time series SIF estimates that are needed in localized settings where plant dynamics must be observed continuously. TOC SIF estimates can be derived from hyperspectral radiance measurements of these systems with much lower uncertainty than from airborne and spaceborne platforms due to higher instrument stability, concurrent irradiance and reference panel measurements and lower atmospheric influence. As a consequence, such estimates have provided important in-situ calibration and validation data for SIF products derived from acquisitions recorded in multiple airborne campaigns (see Sec. 1.4.2). Due to the low spatial resolution their adoption for validation of existing spaceborne SIF products is, however, not widespread (see Sec. 1.4.3). With the advent of new SIF estimates from spatially higher-resolved spaceborne platforms (e.g., FLEX and, as argued in Publication IV, DESIS) these ground based systems, however, may prove to constitute valuable independent validation data for a range of spaceborne products as well.

1.4.2 Airborne Sensor Data

A number of airborne platforms have been equipped with sensors with spectral resolutions allowing for passive SIF retrieval. Most notably, the HyPlant FLUO spectrometer [263] has been operated in yearly campaigns since 2014 as the demonstrator version of the FLEX FLORIS [75] sensor. Other airborne sensors that have been used for SIF retrieval include the micro-hyperspectral imaging sensor [332], the Chlorophyll Fluorescence Imaging Spectrometer (CFIS) [91], the FIREFLY (fluorescence imaging of red and far-red light yield) spectrometer [215] and APEX (Airborne Prism Experiment) [59]. The quality of SIF measurements from airborne platforms has substantially progressed allowing for use in actively researched applications concerned with the previsual monitoring of plant stress (see Section 1.2.4) and in research requiring field-level SIF estimates for phenotyping tasks [194]. Due to large operating costs of such setups installed in airplanes, recently an increased interest has also been placed in sensor systems on UAVs [1].

Airborne SIF retrieval setups have had historically a special importance for the derivation of SIF from data acquired on spaceborne platforms. Instrumental effects, calibration artefacts and unsuitable modelling assumptions have an influence on the

reliability of spaceborne SIF estimates. The retrieval of SIF from airborne hyperspectral data represents an important testing application of SIF retrieval methods and associated calibration procedures that are intended to be operationalized for spaceborne SIF retrieval. Furthermore, airborne SIF estimates are acquired to provide reference data for the validation of spaceborne SIF products following a bottom-up validation principle [199] leveraging reference data at multiple spatial scales. For example, [276] has validated the OCO-2 SIF product with CFIS derived SIF estimates. The use of HyPlant – the airborne demonstrator of the FLEX FLORIS sensor (see Section 1.4.3) – for a dedicated validation campaign of FLEX FLORIS is currently also considered [87].

1.4.3 Spaceborne Sensor Data

Globally distributed SIF estimates have been realized from satellite-based instruments with suitable SR, SNR and SSI [356]. These include the Thermal And Near-infrared Sensor for carbon Observation Fourier Transform Spectrometer (TANSO-FTS) on the Greenhouse gases Observing SaTellite (GOSAT) [114, 143], SCIAMACHY [142], GOME-2 [138, 293], OCO-2 [276] on a free-flying satellite and OCO-3 mounted on the Japanese Experiment Module-Exposed Facility International Space Station [73, 85, 281], TROPOMI [151, 352] and the Atmospheric Carbon dioxide Grating Spectrometer (ACGS) onboard the Chinese Carbon Dioxide Observation Satellite (TanSat) [79, 329]. Following [80], current global SIF products can be grouped by their swath coverage and spectral resolution. While SCIAMACHY and GOME-2 products [140, 153] have provided 0.5° resolutions and global coverage in a few days, GOSAT, OCO-2 and TanSat are narrow-swath imagers with smaller footprints down to a few km^2 with longer repeat cycles. No instrument from which currently global SIF products are derived has been designed for SIF retrieval as primary aim but rather for the quantification of Greenhouse Gas emissions. As a consequence, their spatial resolutions do not allow to study vegetation dynamics below the regional scale. The FLORIS sensor onboard ESA's Earth Explorer FLEX [75, 87] will be the first such instrument with an anticipated spatial resolution of 300 m and repeat cycles ranging from 10 to 24 days. Publication III [31] included in this thesis furthermore addresses for the first time the estimation of SIF at 30 m resolution from the DESIS sensor [158] on the International Space Station.

To leverage the multitude of spectrally high-resolved earth observations for a long term SIF product with improved spatio-temporal coverage recent work has concentrated on harmonization strategies of individual SIF products [209]. A number of works [82, 83, 99, 135, 178, 331, 344] have targeted the spatio-temporal downscaling of coarse SIF products by combining them with high-resolution surface reflectance estimates employing statistical machine learning approaches and physiological constraints [260].

1.4.4 Validation

Methods that derive land-surface products from remote sensing data are evaluated using different strategies in Earth observation. These fall into five broadly defined categories [119, 246]. The products can be validated (i) with direct in-situ point measurements of the target variable, (ii) indirectly by making use of accurate modelling where the the target variable can be inferred from precise in-situ point measurements of auxiliary variables that parameterize the model, (iii) making use of cross-validation with existing products of the same target variable, (iv) by leveraging prior

knowledge about system constraints of the underlying model (e.g. with time-series data) and (v) with synthetic data generated using radiative transfer models.

The validation of remote sensing products with in-situ measurements of the target variable is ultimately the most valuable validation strategy as it represents the most independent assessment of prediction errors [318] and is a prerequisite to provide traceable uncertainty budgets [199]. However, direct validation of airborne and spaceborne SIF products with in-situ point measurements is challenging due multiple factors. The vast difference in measurement footprints, sparse datasets where acquisition times align sufficiently, as well as the misalignment of observational geometries are major difficulties that complicate the creation of ground-truth reference datasets (see Section 1.5.3). As a consequence, validation of airborne and spaceborne SIF products has addressed direct validation only recently with the emergence of increased availability of field and tower-based SIF ground measurements.

Lacking sufficient ground measurements in diverse phenological and observational conditions airborne products can be validated quantitatively with synthetic data [54, 58, 192] presupposing accurate modelling of both the physical process generating the at-sensor radiance as well as the instrument modulation and noise. Qualitative assessments have been used in [58, 314, 333] where the estimated SIF estimates were compared to prior assumptions on the dynamics of covariates (e.g., GPP) and SIF itself. Direct validation with concurrently recorded in-situ field and tower-based SIF estimates [22, 221, 238], as well as Publications I and III, still feature large uncertainties (see Section 1.5.3).

In the spaceborne context, validation with synthetic data represents often the only practical quantitative validation strategy of SIF retrieval methods [303, 356] as in-situ data is sparse and subject to large uncertainties when compared to SIF products with much lower spatial resolutions. To alleviate the current lack of reliable in-situ validation data increased interested is placed in the possibility to install hyperspectral spectrometers on existing covariance (EC) towers to generate reliable and continuous ground SIF reference datasets in various ecosystems [87, 110, 347]. A small set of studies have recently targeted the validation of spaceborne SIF estimates with in-situ reference datasets acquired from EC towers for the first time [80, 81, 336, 337, 347]. Similarly, the use airborne SIF products for validation is being discussed in the FLEX validation preparation [87] as part of a multi-scale validation approach. Dedicated flight campaigns with CFIS and HyPlant have for example been conducted for validation studies of OCO-2 [276, 313] and DESIS (see Publication IV [31]) SIF estimates. However, due to the lack of reliable reference data, SIF estimates from most spaceborne sensors have been ranked qualitatively by their correlation performance with global GPP estimates [139, 260] (OCO-2 [233, 275], TanSat [78], GOME-2, [253, 344], TANSO-FTS onboard GOSAT [90]) even though a linear SIF-GPP scaling is itself not necessarily an expected outcome in all circumstances. Furthermore, a number of cross-validations between independent instruments have been conducted as well (SCIAMACHY vs. GOME-2 [140], OCO-2 vs. GOME-2 [275], TROPOMI vs. OCO-2 and GOME-2 [151], OCO-2 vs. TanSat [328]).

1.5 Challenges in SIF Retrieval

The development of SIF retrieval algorithms is challenging due to several factors connected to the measurement setup and the specific fluorescence signal in hyperspectral remote sensing data. In the following paragraphs the most important challenges to precise retrieval of top-of-canopy SIF are summarized.

1.5.1 Signal Strength

The TOC SIF signal accounts for 1 - 2% of the canopy leaving radiance in the near infrared region [91]. As a consequence, sensor related artefacts such as stray light, dark current related noise, spectral miscalibrations and directional sensitivity impact the fluorescence prediction significantly [8, 35, 185, 203]. Disentangling of the fluorescence signal from the reflected sun-light requires a high signal-to-noise ratio of the sensor and the derived radiance product. Studies regarding suitable SNR, SR and SSI have been conducted with a focus on different sensing systems (e.g., GOSAT and OCO-2 [89], FLEX [75, 87], a general overview is given by [356]) as well as on particular retrieval methods [57, 166]. The advanced requirements on the sensor characteristics have triggered the development of specialized ground-based, airborne and spaceborne sensor systems (see Section 1.4) operated mainly by scientific organizations.

1.5.2 Atmospheric Compensation

Similarly to the demanding requirements on the accuracy of the sensor characterization, SIF retrieval in atmospheric absorption bands relies on exact modelling of the radiative transfer through the atmosphere to correct for molecular and aerosol absorption and scattering [250]. This is in contrast to SIF retrieval based on the infilling of solar Fraunhofer lines that rely on spectral regions that are insensitive to atmospheric effects. Atmospheric contributions to the at-sensor signal impact the SIF retrieval since they lead to misattributions of changes in the oxygen features to SIF. Of particular interest for O_2-B and O_2-A absorption band-based SIF retrievals is the oxygen column integrated over the radiative path length which is mainly determined by the distance between surface and sensor. To a lesser degree the atmospheric water content, the pressure profile, the aerosol model and the specific viewing geometry also influence the retrieval [58, 66, 198, 212, 250]. However, depending on the sensor and observation setup the accurate measurement of atmospheric constituents is often not feasible or too costly such that they must be estimated, particularly in airborne and spaceborne contexts.

The atmospheric contribution to the at-sensor radiance is addressed in SFM and FLD SIF retrieval methods in one of two ways. In the first possibility, a preprocessing step called *atmospheric correction* is conducted with accurate measurements or estimates of atmospheric constituents that allow to normalize at-sensor measurements to surface reflectance [251]. The effect of the additional TOC SIF signal on the effective surface reflectance can then be inverted without atmospheric influences. Secondly, the estimation of atmospheric effects can be integrated in the SIF estimation scheme. In this case, the influence of atmospheric effects is modelled as part of the SIF retrieval formulation such that an initial estimate of the atmospheric composition can be adjusted to fit the observations more closely (e.g., Publications I and III). Due to the computational cost of accurate radiative transfer simulations, these SIF retrieval methods, however, typically do not implement a full joint optimization

of atmospheric constituents, SIF and other relevant quantities such as the surface reflectance, but run a small number of experiments with adjusted atmospheric composition in a process called *transmittance correction*. The possibility of fully joint optimization or training of both atmospheric and surface quantities is, however, subject to research and addressed in the Publications I, III and IV comprised in this thesis [29, 31, 33].

1.5.3 Retrieval Method Evaluation

The validation of airborne and spaceborne products with ground-based point measurements of TOC SIF presents major challenges. In the following, an overview of the main challenges is given.

1. **Spatial mismatch** There is an inherent representation mismatch between in-situ SIF measurements from field and tower-based radiometers of SIF on the one hand and airborne and spaceborne SIF products on the other due to the vast difference in measurement footprints of these systems. The spatial heterogeneity of SIF within large pixels, particularly in a spaceborne context, introduces uncertainty, as a single ground measurement typically does not accurately represent the average SIF value of the entire pixel [80, 353] and assumptions about the spatial distribution of SIF within the pixel have to be made. This issue is particularly pronounced in areas with diverse vegetation types or varying canopy structures. Similarly, in airborne validation settings a representation mismatch can impede straightforward validation analysis due to complex canopy structure that introduces strong signal heterogeneity over the footprints caused (i) by mixing of different plant individuals [35, 128] and (ii) by partial shadowing [20, 156, 334].

2. **Temporal mismatch** SIF is a dynamic quantity that reacts on fast time scales to changes in illumination [130, 258]. Its large sensitivity to natural variation in solar irradiation causes significant discrepancies between SIF estimates taken at different times. In particular, under non-noon, cloudy and hazy conditions auto-correlation time scales of SIF time series can become much shorter due to rapid changes in illumination [46]. Evaluation of SIF products must therefore be conducted with data exhibiting high temporal consistency as was realized in Publications I and III by using in-situ time series data with high temporal sampling rates as validation data and in Publication IV by relying on a novel benchmarking dataset consisting temporally closely matching airborne and spaceborne SIF estimates.

3. **Signal correlation**: SIF is strongly correlated to the near-infrared reflectance of plant leaves as both are driven significantly by the chlorophyll concentration. Depending on the training setup of feature-based SIF estimation techniques, performance validation must therefore take into account spurious uninformative cross-correlations (see Publications II [213] and V [212]) as was done in Publication IV [31] with a *reflectance constrained* performance evaluation.

4. **TOC SIF emission anisotropy**: Similar to vegetation reflectance that may be expressed as bidirectional reflectance factor (BRF) [254], TOC SIF exhibits a non-isotropic bidirectional angular dependency [20, 338, 350]. The TOC SIF signal has a directional dependency (i) with respect to the irradiant PAR that fuels photosynthesis and (ii) with respect to the viewing geometry. This viewing angle dependency is not only preventing a straightforward downscaling

of TOC SIF to leaf-level ChlF [69] (see Section 1.2.3). Importantly, this dependency renders validation of remote sensing SIF with in-situ measurements more challenging as it introduces the requirement to align the prediction with the in-situ measurement either at recording time or in post-processing. The representative mismatch between in-situ labels and SIF estimates is more pronounced in circumstances where the alignment is technically impossible or the post-processing is subject to large uncertainties such as in data from spaceborne platforms where the viewing geometry is set by orbital requirements [141, 350]. Indirect validation studies of SIF post-processing have shown viewing angle corrections to improve validation performance [121, 348, 349].

The directional dependency related to the solar irradiance is primarily driven by the leaf angle distribution (LAD) and gap fraction of the canopy. LAD determines the total leaf area exposed to the irradiant PAR by varying the mean relative angle between the sun and the leaf surface [222]. Equally, the directional dependency is modulated by the gap fraction which determines the amount of (i) light penetration into the canopy (ii) the scattering and reabsorption of ChlF in the canopy and (ii) the contribution of the fluorescence signal from different canopy layers to the canopy leaving radiance (including the amount of soil reflected radiance visible in single pixels [168]). Depending on the canopy structure, the composition of fluorescence contributions may vary significantly in the diurnal course as a consequence [242]. Reabsorption and scattering are also driven by the vegetation reflectance [323] such that the BRF may introduce additional viewing angle dependency. Due to spectrally variable transmission properties of vegetative tissues, this viewing angle dependency is spectrally variable.

1.6 Machine Learning for SIF Retrieval in Hyperspectral Imaging Spectroscopy

1.6.1 Hyperspectral Imaging Spectroscopy

Imaging spectroscopy is a technique in optical remote sensing that captures sensor-incident radiance in narrow spectral bands to sample the electromagnetic spectrum at a high spectral resolution over a spatially contiguous field of view. It provides spatially and spectrally structured information of the interaction between solar irradiation and terrestrial molecules and particles that modulate the measured at-sensor radiance through absorption, scattering and reflection, enabling a synchronous characterization of the Earth's surface and the atmosphere over extended areas [96, 231]. Unlike multispectral sensors that typically collect data in wide spectral bands (e.g., Sentinel-2 [84] and Landsat [289]), hyperspectral sensors acquire spectral information in more narrow bands with spectral resolutions of less than 10 nm, usually covering spectral windows in the spectral range from the visible to the thermal infrared (300 nm - 2500 nm). This high spectral resolution allows for a more precise characterization of surface properties based on finer spectral signatures than is possible with multispectral sensors [41].

Both multispectral and hyperspectral imaging spectroscopy have become a rich sources of physically well characterized imagery in the last decades with airborne sensors such as Aviris-NG [49], APEX [255, 287], HyPlant [263] as well as satellite missions with global coverage such as the Sentinel-2 [84] and Landsat programs [289]. Additionally, in the last years multiple hyperspectral imaging sensor systems

have been operationalized (PRISMA [220], EnMAP [115], DESIS [158], EMIT [283]) or have entered advanced planning stages (FLEX [75], SBG [271], PRISMA2GEN [5]). While the first imaging spectrometers were designed to allow for mineral mapping based on material-specific spectral responses [231], airborne and spaceborne hyperspectral imaging sensors today acquire data for the estimation of environmental quantities without trivial spectral response features such as fire and shadow related or biophysical variables such as the leaf area index (LAI), water quality parameters, and atmospheric constituents to detect subtle variations distributed across large sections of optical reflectance spectra.

1.6.2 Towards Neural Network-based SIF Retrieval

The semantic attribution and disentangling of complex spectral signatures in various remote sensing tasks is typically solved by leveraging Machine Learning methods. The term *Machine Learning* (ML) denotes a field of techniques in artificial intelligence where computing systems are enabled to learn statistical relationships in data to solve tasks without being explicitly programmed to do so. *Learning* in this context describes the capacity to adjust the completion of such a task iteratively under targeted variation of model parameters to improve a mean task performance according to a user defined metric over a *training* dataset. In the context of hyperspectral imaging, *machine learning* targets the identification of generalized spatio-spectral features for tasks such as land use and land cover (LULC) classification [206], object [174] and anomaly [321] detection, mineral abundance mapping [261], mapping of environmental quantities [290, 346] and disaster impacts [236].

Machine Learning methods are adopted in hyperspectral image analysis due to their capacity to effectively distill informative features from large amounts of data despite the large observational variability in spectral data. As in many other fields, ML methods excel over the use of traditional *hand-crafted* features [51]. Traditional ML methods to achieve this include Gaussian Process Regression, Support Vector Machines (SVMs), Decision Trees and Random Forests [136, 224]. Due to the high data dimensionality which the high spectral resolutions of hyperspectral sensors entail the analysis and utilization of hyperspectral of ML methods necessitate the use of dimensionality reduction methods. ML tasks in hyperspectral remote sensing are therefore typically formulated as learning tasks over reduced feature spaces (e.g., [62, 68, 244, 301]) established with custom dimensionality reduction techniques [102, 239, 240]. In recent years, artificial neural network (NN) methods, particularly deep learning architectures, have become increasingly prevalent in hyperspectral applications [118, 177, 206] due to their improved performance in various remote sensing tasks [136, 270] that can be attributed [206] to the well-established improved identification of robust features of DL over traditional ML methods [72, 157, 202]. NN architectures allow to integrate dimensionality reduction and task completion in a single end-to-end prediction scheme. This allows the derivation of more informative features than is possible in traditional ML methods. The increased amount of hyperspectral data sources globally, that benefit the adoption of NN-based methods, as well as the larger computational resources at the disposal of research institutions additionally supports the increased interest in NN architectures. The large amount of hyperspectral data available could be shown to be benefit the training of foundation models in estimation tasks such as land cover classification, segmentation, change detection and unmixing (e.g., for multispectral [126] and hyperspectral [26, 230, 256, 306, 320] sources).

While the recent advances in Deep Learning and in the emerging field of Foundational Modelling in particular have allowed to achieve new state-of-the-art performances in hyperspectral remote sensing [26], their use for SIF retrieval not yet been widely established. Related work includes [139] presenting a supervised approach to post-process noisy SIF estimates to a noise-reduced SIF product with a shallow neural network from the leading Principal Components of hyperspectral GOME-2 observations and independently retrieved SIF estimates. A number of works have furthermore investigated SIF estimation based on broadband surface reflectance products by leveraging supervised neural network training with independent OCO-2 and TROPOMI SIF estimates used as labels. Since such methods do not target the physiological fluorescence signal in the observational data but a non-causal statistical relationship between surface reflectance and SIF, they are, however, not useful as standalone fluorescence estimators, even though they might prove useful for SIF-dependent tasks such as GPP estimation. Recently, the Publications I, III and IV [29, 31, 33] included in this thesis have introduced a neural network-based approach relying on self-supervised training for the first time. As these contributions show good retrieval performances in validation studies with in-situ data and in SIF product cross-validations, their method formulation may lay the basis for future research in neural network-based SIF retrieval.

1.7 Research Objectives

The objectives of this thesis focus on the use of self-supervised Deep Learning for the retrieval of SIF in hyperspectral image data acquired on an airborne and a spaceborne platform. To this end, a sensitivity (Publication V [212]) and a modeling study (Publication II [213]) have been conducted to allow for the development and validation of two SIF prediction methods (Publications I and III) adapted to two representative sensors – DESIS and HyPlant FLUO. While these two sensors feature vastly different specifications, operation modalities and post-processing, they represent different extremes of data quality that can be expected in modern imaging spectrometer equipment suitable for SIF retrieval in airborne and spaceborne contexts. The objectives of this thesis were therefore:

1. Develop a deep learning framework for physically-based SIF retrieval in HyPlant FLUO data around the O_2-A absorption band by adopting key concepts of the well-established optimization-based Spectral Fitting Method (SFM) [52]. Due to the exceptional sensor design of HyPlant FLUO focusing on SIF retrieval in the O_2-A and O_2-B absorption bands the method development could rely on high-fidelity baseline predictions such that method-related biases could be studied effectively in real data (Publications I [33] and III [29]).

2. Contributing to a sensitivity analysis of surface, atmosphere and sensor parameters in typical HyPlant and DESIS observation conditions with respect to the at-sensor fluorescence signal (Publication V [212]). This has allowed the simulation of a large dataset of training spectra that was used to represent the physical generation of the at-sensor radiance signal (Publication II [213]).

3. Improving the physical atmospheric radiative transfer formulation in SFM to allow for local changes in observation geometry (e.g., because of variable topography) without noise-prone preprocessing (Publication III [29]). To this

end, a suitable emulation strategy of atmospheric radiative transfer simulations had to be derived for both DESIS and HyPlant to allow an efficient online simulation of at-sensor radiance during training (Publication II [213]).

4. Provide the first SIF estimates from DESIS acquisitions by adapting the methodology developed and validated on high-quality HyPlant FLUO data (Publication IV [31]).

1.8 Contributions

In this thesis multiple novelties are presented that have the potential to benefit future research in SIF retrieval from airborne and spaceborne hyperspectral imagery. The thesis is based on four cumulative publications. The individual contributions in each publication to research in SIF estimation from hyperspectral spectroscopy data are listed below.

Publication I

1. A feature-based spectral fitting method called Spectral Fitting Method Neural Network (SFMNN) is introduced implementing for the first time a self-supervised deep learning-based based training for SIF retrieval.

2. A loss formulation is proposed that considers physical and physiological constraints as well as a spectrally dependent weighting in the O_2-A band.

3. It is shown that this method can approximately disentangle fluorescence variation from topographic changes in hilly terrain which sets it apart from previous methods for SIF retrieval in airborne contexts.

4. SFMNN reaches competitive validation results on a dataset of in-situ top-of-canopy SIF measurements.

5. HyData [34], a publicly available dataset, is assembled from data recorded in HyPlant campaigns from 2018 to 2023.

Publication II

1. The approximation accuracy of multiple RTM emulation models (polynomial, neural network, kernel based and Gaussian Process regression-based) are tested in the specific context of SIF retrieval from at-sensor radiance in the O_2-A absorption band acquired by HyPlant FLUO and DESIS by drawing on an exhaustive sensitivity analysis (Publication V [212]).

2. A fourth-order polynomial emulation model based on a fixed feature space consisting of the power set of the parameterizing variables is shown to be sufficiently accurate, fast and robust to the training dataset size. These qualities have the potential to allow to speed up and widen the application of accurate RTM models in various remote sensing applications.

Publication III

1. The emulator of at-sensor radiance in the O_2-A band developed in Publication II is extended to allow SFMNN (Publication I) to estimate atmospheric constituents and sensor miscalibrations simultaneously with SIF.

2. The combined use of this emulator and the loss formulation of SFMNN are shown to allow the creation of a state-of-the-art SIF retrieval method for HyPlant FLUO data. In particular, the prediction performance is improved over the model presented in Publication I.

Publication IV

1. The first SIF retrieval method for DESIS acquisitions is introduced and validated with a validation dataset [32] consisting of quasi-simultaneous high-fidelity HyPlant-derived SIF estimates as well as with OCO-3 observations.

2. The loss developed in Publications I and III is extended to suit low-resolution DESIS data and shown to improve validation performance. (i) A perturbation-based regularizer and (ii) supervised loss terms targeting atmospheric estimates are introduced.

The methods and concepts developed in this thesis are generally tested both in DESIS and HyPlant FLUO such that adaption to different sensors with similar characterizations are likely to succeed with similar estimation errors.

Chapter 2

Summary of Publications

This thesis is based on four publications. In this chapter, a short summary of all publications is given by highlighting the underlying research questions, methodology and validation results. The publications are attached as separate Chapters A - D.

2.1 Publication I

In the last decade, the data acquisition for the derivation of spatially highly resolved SIF maps from airborne platforms has become achievable. The operational start of ESA's Earth Explorer mission FLEX will conclude this development also for spaceborne platforms. On the one hand, this development was triggered by advances in hyperspectral sensor design enabling to develop sensor systems capable of acquiring hyperspectral image data from airborne and spaceborne platforms with SNRs suitable for the retrieval of sun-induced fluorescence (SIF). On the other hand, the targeting of oxygen absorption bands in spectral fitting and Fraunhofer discrimination methods has allowed to derive fluorescence at lower spectral resolution and SNR than required in methods targeting very fine solar absorption features. This has allowed to improve the spatial resolution of these systems. The high spatial resolution of such SIF products is notably projected to be beneficial for environmental research and monitoring due to the close causal link between fluorescence and carbon and water fluxes in vegetation canopies.

Publication I [33] introduces a novel approach for SIF retrieval relying on a Spectral Fitting Method Neural Network (SFMNN). SFMNN is a neural network-based self-supervised method that addresses key challenges faced by traditional SIF retrieval algorithms. Well-established methods like 3FLD, iFLD, and the Spectral Fitting Method (SFM) decompose at-sensor radiance into fluorescence and reflectance components by relying on pixelwise optimization assuming constant atmospheric conditions for efficiency reasons. SFMNN is inherently a feature-based method such that it has the potential to alleviate the need for recurring optimization. Furthermore, SFMNN adapts to local changes of the atmospheric radiative transfer as occurs for example in hilly terrain where the optical path length varies significantly. This is implemented without needing separate atmospheric correction steps, making it particularly promising for topographically complex regions. In summary, key technical novelties of SFMNN include:

1. Self-supervised loss function: Unlike supervised neural network methods for SIF retrieval that rely on labeled data for training, SFMNN uses a self-supervised loss function to invert a physical model of the at-sensor radiance. This approach does not require manually annotated ground truth data for SIF, improving its flexibility and scalability.

2. Joint estimation of fluorescence, reflectance, atmospheric transfer and sensor miscalibration: The method jointly estimates surface reflectance, fluorescence emission and the influence of atmospheric transmittance and scattering during data acquisition. This contrasts with baseline methods that handle these components separately or rely on fixed assumptions for atmospheric conditions.

3. Adaptability to local atmospheric conditions: One of the most significant advancements of SFMNN relies in its capability to learn and adjust to local atmospheric variations. The method does not assume constant atmospheric conditions across the scene, which is a major limitation of existing methods. By learning atmospheric transfer functions as part of the retrieval process, it provides more accurate and localized SIF estimates, especially in regions with varying topography and atmospheric properties.

4. No need for separate atmospheric correction: Traditional retrieval methods require atmospheric corrections or localized atmospheric characterizations, which can introduce errors or require significant computational overhead. In contrast, SFMNN integrates the estimation of atmospheric transfer directly into the model, avoiding these additional steps and reducing computational costs.

Publication I presents a validation study of SFMNN with traditional methods with the benchmark dataset HyData [34]. This dataset combines HyPlant acquisitions and in-situ top-of-canopy SIF measurements. While SFMNN does not outperform all baseline methods in this validation dataset in terms of accuracy due to an overestimating bias, its correlation with respect to in-situ measurements outperforms baseline methods in a majority of validation sets. Furthermore, two qualitative application studies highlight the physiological and physical consistency of SFMNN.

In summary, SFMNN introduces a powerful, efficient and adaptive approach for SIF retrieval, addressing the limitations of traditional methods by jointly estimating key parameters. Its ability to adapt to local conditions and handle complex terrain makes it a promising tool for advancing remote sensing of plant health and productivity.

2.2 Publication II

SIF retrieval relies on accurate modelling of sensor-related artefacts, illumination and viewing-related geometry as well as atmospheric effects. These effects can have an imprint on the measured radiance of the hyperspectral imaging sensors HyPlant FLUO and DESIS on the same order of magnitude as SIF itself (Publication V [212]). Typical approaches adopted in various SIF retrieval methods leverage accurate Radiative Transfer Models (RTMs) to account for geometry and atmospheric-related effects and separate spectral corrections to account for sensor artefacts in dedicated preprocessing steps. In practice, the parametrization of RTMs can not be derived from measurements alone since especially the atmospheric constitution is rarely known at acquisition time. The computational cost of accurate RTMs, however, prohibits the estimation of undetermined parameters during SIF retrieval by direct integration of the RTM into the learning or optimization process of the SIF retrieval methods. RTM emulation provides a computationally efficient approximation of an RTM model, reducing the computational burden while maintaining physical accuracy at a sufficient level. Publication II therefore investigates the accuracy and speed of RTM emulation based on polynomial, neural network, kernel ridge and Gaussian

Process (GP) regression in the specific context of SIF retrieval in the O_2-A absorption band of the hyperspectral HyPlant FLUO and DESIS imaging sensors.

Publication II [213] finds excellent emulation performance within the small spectral window around the O_2-A absorption band with a fourth-order polynomial over a fixed feature space. The approximation residual of this emulator with respect to the RTM can be shown to be sufficient for SIF retrieval while also exhibiting low prediction times. Its performance is notably better than the kernel-based, GP-based and neural network-based setups that were equally tested. Publication II argues, that while improved results might be gained from these methods after hyperparameter tuning, the adoption of a polynomial emulator in SIF retrieval methods still is preferable in many use cases due to its fast prediction time, performance robustness under small training dataset sizes and model simplicity that, notably, allows a straightforward inclusion in the gradient backpropagation mechanism in neural network training.

With Publication II a simple yet accurate polynomial RTM emulation method the O_2-A absorption band is presented to model HyPlant FLUO and DESIS at-sensor radiances under typical viewing geometries and atmospheric conditions. Since its accuracy is sufficient to disentangle SIF from other effects on the measured the at-sensor radiance, it qualifies for use in machine learning-based SIF retrieval methods in HyPlant and DESIS data. However, the application of this approach is not restricted to these two sensors as it can be transposed to other hyperspectral sensors with suitable spectral configurations such as, e.g., FLEX [75], and possibly other spectral ranges such as the O_2-B thereby extending the application of machine learning methods for SIF retrieval.

2.3 Publication III

Publication III [29] focuses on improving the SFMNN methodology outlined in Publication I by enhancing the physical accuracy of its at-sensor radiance simulation. Since SFMNN targets the reconstruction of the atmospheric oxygen absorption line O_2-A, it requires exact modelling of the down and upwelling radiative transfer through the atmosphere. The atmospheric state at acquisition time is most commonly undetermined such that the atmospheric characterization must be estimated as part of the SIF retrieval. In SFMNN the atmospheric transfer functions are estimated as interpolations in a PCA space spanned by observed transfer functions derived from a subset of all HyPlant acquisitions. However, such a formulation can not guarantee to produce physically plausible transfer functions. Furthermore, it does not establish an explicit relationship between the physical drivers and the functional form of the transfer function.

Atmospheric radiative transfer models (RTMs) such as MODTRAN6 [25], 6S/6SV [154, 155] and libRadTran [86] are typically used to model precisely atmospheric transfer functions, but their high computational cost makes them impractical for pixel-wise SIF retrieval or direct use in the gradient backpropagation of neural network training. To alleviate this inefficiency, a fast and accurate indirect use of RTMs is proposed to replace the implicit atmospheric formulation adopted in SFMNN. Specifically, Publication III introduces an emulator-based version of SFMNN (Em-SFMNN), which integrates the polynomial RTM emulator developed in Publication II [213] and [211] with the basic SFMNN concepts outlined in Publication I. Additionally, a simple, but computationally efficient extension to this polynomial emulator

formulation is introduced in order to improve the representation of spectral sensor miscalibrations.

Publication III [29] includes a quantitative validation study of its SIF estimation performance and two qualitative case studies to assess its consistency across different datasets and of auxiliary estimates. In the comparative validation study EmSFMNN is validated using HyPlant campaign data and in-situ TOC SIF measurements. EmSFMNN is shown to demonstrate strong agreement with ground-based in-situ SIF measurements outperforming both SFMNN and traditional baseline methods in most of the validation datasets. Furthermore, the study shows that a backbone model without finetuning is able to yield SIF predictions with validation errors similar to traditional methods.

In addition, Publication III studies qualitatively its SIF estimation in HyPlant acquisitions with strong topographic gradients. In traditional methods additional atmospheric correction methods steps are needed to allow SIF retrieval uniformly in such acquisitions in order to account for differences in the optical path length of the radiative transmission trough the atmosphere and the boundary layer. Publication III shows that EmSFMNN yields consistent SIF estimates across topographical gradients without any further processing steps, highlighting that the implicit atmospheric correction of the RTM emulation layer is consistently integrated in the prediction scheme.

Publication III furthermore presents a qualitative validation of the EmSFMNN's prediction of aerosol optical thickness (AOT). To this end EmSFMNN's mean AOT predictions in a multi-day time series of 13 HyPlant FLUO acquisitions were compared (i) to high-quality AOT estimates derived from sun photometer measurements recorded at the FZJ-JOYCE AERONET station [4] and (ii) with MODIS derived AOT products used as ground truth. This comparison could show consistent dynamics in EmSFMNN-predicted AOT and the ground truth sun-photometer measurements. This result further underscores that the targeted disentangling of atmospheric effects, surface reflectance characteristics and SIF is indeed realized by means of the constraint-based loss of EmSFMNN.

In Publication III a retrieval method is developed that allows for the first time consistent SIF estimates over variable topography without noise-prone preprocessing steps. Its good validation performance with respect to high-quality in-situ TOC SIF measurements shows that finetuning a common baseline model yields state-of-the-art SIF predictions from HyPlant FLUO. Since HyPlant FLUO is the demonstrator sensor for FLEX, a future application of this approach to this spaceborne sensor with minimal sensor-related domain gap provides an interesting opportunity to test this approach in a global scenario.

2.4 Publication IV

DESIS is a hyperspectral sensor onboard the International Space Station (ISS) with a significantly lower spectral resolution than HyPlant FLUO. Its design specifications were not chosen to fulfill optimal SR, SNR and SSI requirements for traditional SIF retrieval methods as for HyPlant FLUO. As a consequence, SIF retrieval from DESIS has never been addressed in the past even though its high spatial resolution of 30 m is makes it an interesting sensor for field-scale predictions of environmental variables. In Publication IV [31], the EmSFMNN architecture and loss, that have been originally developed for HyPlant FLUO, are adapted to SIF retrieval in DESIS

data. The adapted EmSFMNN methodology is shown to allow for the first time to derive SIF estimates at 30 m resolution from data acquired by a spaceborne sensor. Currently, spaceborne SIF estimates are derived from atmospheric missions such as GOSAT, OCO-2/3, and TROPOMI, which have sufficient spectral resolution but low spatial resolution (>4 km^2). ESA's Earth Explorer Mission FLEX, scheduled for launch in 2026, will provide estimates with 300-meter resolution, offering significant improvements but still falling short for many applications relying on higher spatial resolutions. The proposed EmSFMNN implementation has therefore the potential to provide a valuable addition to a range of SIF products where high-resolution SIF imagery is needed such as in precision agriculture. Furthermore, this DESIS SIF product may complement the validation efforts of FLEX SIF product by providing cross-comparison estimates.

In Publication IV several technical innovations to EmSFMNN have been implemented in order to allow SIF estimation from DESIS. Since the spectral resolution of DESIS is much smaller than the resolution of HyPlant FLUO (2.5 nm vs. 0.25 nm) the changes to the original EmSFMNN methodology aim at constraining the training of the spectral reconstruction more strongly. Key adaptations in the approach include a perturbation-based augmentation scheme to minimize the influence of confounding variables, the integration of atmospheric data from DESIS Level 2A products through supervised learning tasks as well as the use of a supervised SIF predictor trained on simulation data.

Publication IV includes a validation study of the newly developed DESIS SIF product with high-quality SIF estimates derived from spatially matching and quasi-synchronous HyPlant acquisitions in North-Western Germany [**?**]. The acquisitions of HyPlant FLUO were scheduled specifically to allow for a this validation with a time difference between the SIF products of less than 20 minutes to reduce the signal decorrelation due to the dynamic nature of SIF. The validation shows close alignment in the two SIF products of 0.78 mW nm^{-1} sr^{-1} m^{-2} at 740 nm and a coefficient of determination $r^2 = 0.6$. Furthermore Publication IV also compares this novel DESIS SIF product to globally distributed OCO-3 SIF estimates. Since OCO-3 is equally located a dataset of the ISS matching acquisitions within small time windows of less than 10 minutes could be established. Despite large differences in the spatial resolution between the two sensors, the validation still finds a weak correlation between the SIF products ($r^2 = 0.2$).

Publication IV shows that EmSFMNN can be extended to spaceborne sensors for SIF estimates with reasonable uncertainties even in the case of very challenging sensor characteristics. As a consequence, the EmSFMNN SIF retrieval methodology can be expected to be applied successfully to future hyperspectral missions such as FLEX that feature sensor characteristics adapted to SIF retrieval.

Chapter 3

Discussion

3.1 Self-Supervised Training for SIF Estimation

Self-supervised training adopts a loss formulation that does not rely on labels but on intrinsic data properties or assumptions about the data generation process to guide the learning process [262]. It is an important learning framework in various fields with Deep Learning applications, for example in natural language processing (NLP) and computer vision. In NLP, it has been instrumental in the development of foundation models like BERT [70] and GPT [201], which use techniques such as masked language modeling and autoregression [132]. In computer vision, self-supervised learning has been adapted by means of reconstruction-based and contrastive techniques [131]. A typical reconstruction-based formulation in self-supervised learning are Masked Autoencoders (MAEs). MAEs work by masking random patches of input images and tasking the model with reconstructing the missing pixels. In contrastive techniques, equivariance and invariance properties of the data or the desired predictor are leveraged to formulate a training scheme.

Variations of such reconstruction-based and contrastive learning formulations from general computer vision have equally been implemented for network training in remote sensing data to contend with its specific features and requirements such as multi-modality [15, 95, 230, 320], large spectral input dimensionality [55, 126, 190], seasonality and temporal variation [11, 184, 288], observation angle dependencies [71, 164] and specific noise patterns [161]. Self-supervised training is beneficial in contexts where labelled data is costly or unavailable such as in SIF estimation explored in this thesis. Accordingly, this thesis has applied a self-supervised reconstruction-based methodology of traditional spectral fitting methods to train the predictor neural network in Publications I and III and, partly, in Publication IV. This has been achieved by relying on the use of an auto-encoder and two different physical radiative transfer model formulations. While in Publication I linear combinations of Principal Components over a representative set of radiative transfer functions were used (see Publication I, Sec. 3.1.), SIF prediction performance in Publication III could be improved by leveraging an accurate emulation framework developed in Publication II (see Publication III, Sec. 3). Both physical model formulations are integrated directly into a reconstruction-based loss to be able to conduct the estimator training with measured data, similarly to [124] where a simulation based on estimated quantities is compared to the input. Such a procedure can be qualified *self-supervised* since implicit properties of the physical radiative transfer are leveraged to formulate the training loss instead of reference data serving as labels. In addition, a self-supervised regularization method has been developed and validated as part of the EmSFMNN training on DESIS data in Publication IV to leverage a physical invariance property in a perturbation based loss regularizer (see Publication IV, Sec. 3.2.). The self-supervised approaches for SIF retrieval followed in this thesis have

allowed to include domain-specific knowledge about the physical signal generation process.

3.2 Coupling Hyperspectral Emulation and Deep Learning

Estimation of SIF based on hyperspectral data in the oxygen absorption bands must involve atmospheric correction steps prior to the retrieval process or an atmospheric characterization as part of the retrieval method. This accounts for the oxygen absorption as well as modulating effects of the atmospheric transfer driven by the water vapour and aerosols present in the optical path. The sensitivity of the at-sensor radiance in the O_2-A absorption band to optical path length is such that in SIF retrieval atmospheric correction has to be considered for small surface-sensor distances in the range of a few meters already [252]. However, atmospheric correction steps represent a significant amount of the total processing time of traditional SIF retrieval methods. As a relief to this problem, a feature-based approach for SIF estimation has been adopted throughout Publications I, III and IV where the pixelwise optimization of traditional SFM [52] is replaced by neural network inference. This allows costly atmospheric correction preprocessing steps to be substituted by fast online estimation of the atmospheric effects. To this end a neural network is trained with a simulation-based loss that reconstructs spectra from disentangled physical quantities. The influence of the atmospheric radiative transfer is addressed by specifically adopting a parameterization of the reconstruction simulation that allows for variation in atmospheric quantities. While the consistency of the atmospheric estimates in Publication I has suffered from inconsistent radiative transfer approximations (see Publication I, Sec. 5.1.), the preparatory work of Publication II improved the simulation during training. In this work, an emulator acting as surrogate model for a precise but computationally demanding RTM [212] was used, similarly to well-known look-up table based approaches for the retrieval of surface quantities from remote sensing data [257]. These equally rely on large scale preliminary simulation over a dense sampling set of the relevant surface and atmospheric parameters to derive a functional approximation. Differently to LUT applications, however, the derived emulator was not implemented as an interpolator over the set of precomputed simulations but as pretrained polynomial to allow for efficient batchwise simulations.

As discussed in Publication III, the specific emulation formulation developed in this thesis has allowed a joint reflectance and SIF retrieval, atmospheric correction and estimation of the instrument characterization similarly to Optimal Estimation (OE) techniques [282, 284, 285], where the atmospheric correction is implemented as a loss minimization during training. Notably, integration of RTM emulation approaches in OE have been proposed and validated for surface reflectance estimation from hyperspectral imagery with a neural network emulator on PRISMA [28] and with a hybrid emulation technique on Aviris data [27]. In this thesis, the beneficial use of RTM emulation in the context of SIF retrieval has, however, been shown for the first time.

3.3 Validation Accuracy

In this thesis, the coupling of physical radiative transfer models with neural network training could be shown to yield validation results with state-of-the-art validation results comparing to SFM [52] and iFLD [9] baseline methods (see Publication I,

Tabs. 6 and 7, and Publication III, Tab. 5). In particular, the improved physical precision and range of control in the emulator formulation in Publication III has lead to validation errors of SIF at 760 nm in the range 0.2 - 0.4 mW nm^{-1} sr^{-1} m^{-2} with a dataset of in-situ measurements acquired in multiple years in tandem with airborne HyPlant campaigns. The extension of this approach and application to DESIS data in Publication IV yielded a mean error of 0.78 mW nm^{-1} sr^{-1} m^{-2} at 740 nm (corresponding to $\sim$ 0.36 mW nm^{-1} sr^{-1} m^{-2} at 760 nm) with respect to an SFM baseline in a benchmark set acquired specifically for this validation study. In a comparison with globally distributed OCO-3 soundings, the DESIS SIF product performed worse due to strong overestimation (see Publication IV, Fig. 5). While validation metrics did not meet the challenging FLEX mission objectives to provide SIF at 760 nm with an accuracy of 0.2 mW nm^{-1} sr^{-1} m^{-2}, it must be noted that both in the HyPlant (Publication III) and DESIS (Publication IV) prediction major additional sources of uncertainty were present that will be reduced in FLEX' specific measurement setup: (i) first, differently to the HyPlant prediction case, synchronous retrievals of AOT and water vapour density from Sentinel-3 will be able to be leveraged to constrain the radiative transfer similarly to Publication IV, where a more detailed parameterization is shown to improve the SIF retrieval accuracy (see Publication IV, Sec. 4.3.). (ii) Secondly, FLEX acquisitions will exhibit spectral resolutions that are very close to HyPlant FLUO, thus providing significantly better suited data for SIF retrieval than the test of the proposed methodology on spaceborne data (DESIS) conducted in Publication IV.

Quantifying the accuracy of SIF estimates is a challenging task that has been conduycted in this thesis with in-situ datasets [34] and cross-platforms comparisons [32]. Major sources of uncertainty regarding the validation could, however, not be addressed. The used in-situ data sets were not collected with validation of a SIF product in mind, such that emission anisotropy, viewing angle mismatch, performance variation in different vegetation types and the impact of pressure and temperature profile variation could not be considered in the analysis (see Sec. 1.5.3). These factors contribute, however, to the uncertainty in SIF estimates and their validation such that further efforts must be made to create a well-defined validation setup for SIF retrieval method validation. In this context, the quasi-simultaneous data set consisting of HyPlant and DESIS acquisitions [32] represents an exceptionally useful validation benchmarking data set due to the small time intervals between the recording of the two sensors despite non-optimal viewing angle differences in spaceborne and airborne acquisitions. Similar data quasi-simultaneous acquisition from airborne and spaceborne platforms for SIF retrieval validation have previously only been performed for the validation of OCO-2 with CFIS [276] and they are considered for the future validation of FLEX FLORIS SIF products [87]. An extended interest in the data set and repeated quasi-simultaneous data acquisition with DESIS in future campaigns may provide a sufficiently diverse validation data set of DESIS to allow it to be used for the cross-validation of the future FLEX SIF products. Orbit unpredictability of the ISS, however, makes a careful planning of the alignment of HyPlant and DESIS for a repeated simultaneous data acquisition difficult. Furthermore, at the time of writing it is not clear if the operational life time of DESIS will be long enough to allow for a time window of simultaneous DESIS and FLEX operation.

3.4 Use of Non-Physiological Models for SIF Estimation

Much progress has been made in the mathematical description of molecular and leaf-level SIF [109, 129, 278, 292, 338]. The simulation formulations for surface reflectance and TOC SIF emission adopted in this thesis do not follow the most advanced models describing these quantities in vegetation canopies (e.g., SCOPE [292, 300]). This has two main reasons. First, since the reconstruction-based loss only targets a small spectral range around the O_2-A absorption band (740 - 780 nm), it is not necessary to cover much of the complexity expressed in state-of-the-art physiological models that cover much larger spectral domains. For example, the Gaussian approximation of one of the two modes in the chlorophyll fluorescence emission is valid to a high degree in the chosen spectral range [212]. The reflectance model on the other hand has been developed and validated on an extensive set of DUAL derived surface reflectances [213]. Its parameterization has the beneficial property that it can be monotonically sampled by iterating over linearly defined ranges. However, the reflectance validation data set lacks representativeness outside the specific land cover types over which HyPlant is operated, i.e. healthy forests and crops in mid-latitude summer conditions, and does not represent well human-made structures, as well as water surfaces such that it may be necessary to generalize the reflectance modelling in future extensions of the EmSFMNN approach (Publication III) to SIF prediction from general observation modalities.

A second feature of this simplistic surface model is given by its lack of intrinsic causal structure. The modelled reflectance and TOC fluorescence are not related via the solar irradiance, canopy structure, plant physiology, surface reflectance and viewing geometry. Thus, while the tight integration of physical simulation and emulation in the reconstruction-based loss terms in Publications I, III and IV impose strong implicit constraints on the training, the feature-based prediction of physiology related quantities is guided by basic assumptions on spectral reflectance and fluorescence emission shapes. In this thesis' learning-based approach this prevents the unintended biasing of the SIF estimator to other fitted parameters such as the chlorophyll content that both drives NIR reflectance and fluorescence in physiological models.

Chapter 4

Outlook

4.1 Towards SIF Prediction from Improved Encoding of Hyperspectral Data

Foundation models act as powerful task-agnostic encoders capable of representing diverse data types. As a consequence, these models can be adapted to various applications, enhancing performance in many computer vision tasks significantly. Remote sensing related applications have benefited from innovations in this field for classification, segmentation and mapping (regression) tasks [16, 319]. The use of large foundation models has proven to improve significantly the performance of neural network models in a wide range of vision related tasks [26, 230, 256, 306, 320]. The similarity of multi-band data sources to classical imaging has allowed for straightforward implementation of concepts originally developed in natural images. Applications on hyperspectral data as used in Publications I - III of this thesis usually require some model adaptation due to the larger memory consumption for single observations [230]. The availability of large datasets for foundation model training is crucial [7, 146, 340] as it must scale with the large model size to allow generalization of the feature space. This generalization capability makes foundation models particularly valuable in scenarios where task-specific data may be limited, as they can be adapt to new downstream tasks on smaller data set sizes [235]. With the advent of a sizeable number of continuously recording spaceborne hyperspectral imaging sensors such as PRISMA [220], EnMAP [115], DESIS [158], EMIT [283] as well as hyperspectral pointing instruments such as TANSO-FTS [114, 143], SCIAMACHY [142], GOME-2 [138, 293], OCO-2 [276] and OCO-3 [73, 85, 281], TROPOMI [151, 352] and TanSat [79, 329] the available data set sizes qualify for foundational training. While the use of benchmarking data sets for hyperspectral training is still not as widespread [26] as in the multi-spectral domain [126, 159, 272] domain first contributions in this domain have addressed the training of foundational models on data sets acquired by individual [26] hyperspectral sensors. A key strength of foundation models is, however, their ability to derive embeddings from multi-modal input. Research on multi-modality for sensor fusion in the context of hyperspectral data has been driven by the goal to make use of larger simultaneous observational coverage and to leverage complementary sensor qualities in repeated or (quasi-) simultaneous observations [230, 320].

In Publications I, III and IV, included in this thesis, a small neural network architecture has been chosen due to its versatile feature identification properties. With a self-supervised loss formulation it has allowed to leverage prior knowledge about the physical generation process of the observational data making use of intrinsic constraints encoded in the tightly integrated radiative transfer and instrument model formulations. The proposed training schemes have adopted concepts from classical self-supervised training for computer vision tasks. While a fully general model for

SIF estimation in HyPlant data could be shown to reach the performance of baseline methods, significantly better results were achieved after finetuning to individual campaign data sets (Publication II). Furthermore, the training data sizes for HyPlant and DESIS data were limited due to a constrained set of available data (HyPlant) or available acquisitions that corresponded to the parametrization bounds of the implemented physical models.

An interesting direction for further research on deep learning based SIF retrieval, thus, includes the use of large foundational models as auto-encoders used in the EmSFMNN prediction scheme presented in this thesis. Such models could be pretrained in a task-agnostic way on data that is not suitable for SIF retrieval a priori, but which allows the to acquire feature representations of the atmospheric radiative transfer, typical scene and spectral structures in a larger data base. Preferably such a model should be trained on a range of sensors with different spectral resolutions and overlapping coverage for an improved representation of sensor related effects in contrast to changes in surface related variables such as SIF potentially reducing systematic shifts in existing global SIF products due to orbit and viewing angle differences [141].

4.2 Relevance for SIF Retrieval from FLEX Imagery

Two sensors with significant differences spectral characterizations and spatial resolution were studied in this thesis. These two sensors have exemplified the use of variants of SFMNN for SIF estimation under different data modalities of hyperspectral data sets that are currently and will be available in the near future. In particular, it has allowed to gauge the possibility of applying EmSFMNN to data acquired by the FLORIS sensor onboard the FLEX satellite which is due to be operational in 2026. On the one hand, the HyPlant FLUO sensor is the airborne demonstrator for FLORIS. Challenges in EmSFMNN due to the high spectral resolution could be tested before FLORIS' official launch and data acquisition with observational data. On the other hand, this thesis' work regarding SIF estimation from DESIS data is relevant to FLEX SIF retrieval as EmSFMNN could be tested on observational spaceborne data. While more exact SIF products can be estimated from atmospheric sensors such as from TANSO-FTS or GOME-2, DESIS allows to work with spatially explicit data at high resolution (30 m). Furthermore, the application of EmSFMNN to DESIS in addition to HyPlant has enabled insights into (i) challenges to SIF retrieval specifically related to sensors with low-spectral resolution and (ii) FLEX' tandem set-up with Sentinel-3 for high-quality estimates of atmospheric components. Namely, in Publication III two strategies to include atmospheric estimates into the EmSFMNN training were tested: (i) as a direct parameterization of the simulation layer and (ii) via the inclusion of supervised loss terms.

Finally, EmSFMNN may be relevant for the FLEX mission as a provider of valuable source of cross-validation SIF estimations in the validation and calibration campaign of the FLEX satellite. The EmSFMNN DESIS SIF product is the spatially highest resolved spaceborne SIF product with a pixel size that is significantly smaller than the FLEX pixel (300 m). It may thus complement in-situ validation data gained either from ground-based installations or drone-based products in circumstances where such high-fidelity validation data can not be gathered. Since the operational timeline of DESIS is not fixed at the time of writing this thesis, it is, however, not known whether there will be a time window where both FLEX and DESIS are operated simultaneously.

Bibliography

[1] Aasen, H., Van Wittenberghe, S., Sabater Medina, N., Damm, A., Goulas, Y., Wieneke, S., Hueni, A., Malenovský, Z., Alonso, L., Pacheco-Labrador, J., Cendrero-Mateo, M.P., Tomelleri, E., Burkart, A., Cogliati, S., Rascher, U., Mac Arthur, A., 2019. Sun-Induced Chlorophyll Fluorescence II: Review of Passive Measurement Setups, Protocols, and Their Application at the Leaf to Canopy Level. Remote Sensing 11, 927. doi:10.3390/rs11080927.

[2] Ač, A., Malenovský, Z., Olejníčková, J., Gallé, A., Rascher, U., Mohammed, G., 2015. Meta-analysis assessing potential of steady-state chlorophyll fluorescence for remote sensing detection of plant water, temperature and nitrogen stress. Remote Sensing of Environment 168, 420–436. doi:10.1016/j.rse.2015.07.022.

[3] Acebron, K., Matsubara, S., Jedmowski, C., Emin, D., Muller, O., Rascher, U., 2021. Diurnal dynamics of nonphotochemical quenching in Arabidopsis npq mutants assessed by solar-induced fluorescence and reflectance measurements in the field. New Phytologist 229, 2104–2119. doi:10.1111/nph.16984.

[4] AERONET FZJ-JOYCE, 2024. AERONET Site Information Database. URL: https://aeronet.gsfc.nasa.gov/new_web/photo_db_v3/FZJ-JOYCE.html.

[5] Agenzia Spaziale Italiana (ASI), . PRISMA2GEN - eoPortal. https://www.eoportal.org/satellite-missions/prisma2gen.

[6] Ahmed, K.R., Paul-Limoges, E., Rascher, U., Hanus, J., Miglietta, F., Colombo, R., Peressotti, A., Genangeli, A., Damm, A., 2023. Empirical insights on the use of sun-induced chlorophyll fluorescence to estimate short-term changes in crop transpiration under controlled water limitation. ISPRS Journal of Photogrammetry and Remote Sensing 203, 71–85. doi:10.1016/j.isprsjprs.2023.07.016.

[7] Alabdulmohsin, I., Zhai, X., Kolesnikov, A., Beyer, L., 2024. Getting ViT in Shape: Scaling Laws for Compute-Optimal Model Design. doi:10.48550/arXiv.2305.13035, arXiv:2305.13035.

[8] Albert, L.P., Cushman, K.C., Zong, Y., Allen, D.W., Alonso, L., Kellner, J.R., 2023. Sensitivity of solar-induced fluorescence to spectral stray light in high resolution imaging spectroscopy. Remote Sensing of Environment 285, 113313. doi:10.1016/j.rse.2022.113313.

[9] Alonso, L., Gomez-Chova, L., Vila-Frances, J., Amoros-Lopez, J., Guanter, L., Calpe, J., Moreno, J., 2008. Improved Fraunhofer Line Discrimination Method for Vegetation Fluorescence Quantification. IEEE Geoscience and Remote Sensing Letters 5, 620–624. doi:10.1109/LGRS.2008.2001180.

[10] Amoros-Lopez, J., Gomez-Chova, L., Vila-Frances, J., Alonso, L., Calpe, J., Moreno, J., and del Valle-Tascon, S., 2008. Evaluation of remote sensing of

vegetation fluorescence by the analysis of diurnal cycles. International Journal of Remote Sensing 29, 5423–5436. doi:10.1080/01431160802036391.

[11] An, X., He, W., Zou, J., Yang, G., Zhang, H., 2024. Pretrain a Remote Sensing Foundation Model by Promoting Intra-Instance Similarity. IEEE Transactions on Geoscience and Remote Sensing 62, 1–15. doi:10.1109/TGRS.2024.3469568.

[12] Ananyev, G., Kolber, Z.S., Klimov, D., Falkowski, P.G., Berry, J.A., Rascher, U., Martin, R., Osmond, B., 2005. Remote sensing of heterogeneity in photosynthetic efficiency, electron transport and dissipation of excess light in Populus deltoides stands under ambient and elevated CO2 concentrations, and in a tropical forest canopy, using a new laser-induced fluorescence transient device. Global Change Biology 11, 1195–1206. doi:10.1111/j.1365-2486.2005.00988.x.

[13] Angelopoulou, T., Chabrillat, S., Pignatti, S., Milewski, R., Karyotis, K., Brell, M., Ruhtz, T., Bochtis, D., Zalidis, G., 2023. Evaluation of Airborne HySpex and Spaceborne PRISMA Hyperspectral Remote Sensing Data for Soil Organic Matter and Carbonates Estimation. Remote Sensing 15, 1106. doi:10.3390/rs15041106.

[14] Antala, M., Rastogi, A., Cogliati, S., Stróżecki, M., Colombo, R., Juszczak, R., 2024. Sun-induced fluorescence spectrum as a tool for assessing peatland vegetation productivity in the framework of warming and reduced precipitation experiment. Remote Sensing of Environment 301, 113921. doi:10.1016/j.rse.2023.113921.

[15] Astruc, G., Gonthier, N., Mallet, C., Landrieu, L., 2024. OmniSat: Self-Supervised Modality Fusion for Earth Observation. arXiv:2404.08351.

[16] Awais, M., Naseer, M., Khan, S., Anwer, R.M., Cholakkal, H., Shah, M., Yang, M.H., Khan, F.S., 2025. Foundation Models Defining a New Era in Vision: A Survey and Outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence , 1–20doi:10.1109/TPAMI.2024.3506283.

[17] Bachmann, M., Makarau, A., Segl, K., Richter, R., 2015. Estimating the Influence of Spectral and Radiometric Calibration Uncertainties on EnMAP Data Products—Examples for Ground Reflectance Retrieval and Vegetation Indices. Remote Sensing 7, 10689–10714. doi:10.3390/rs70810689.

[18] Bacour, C., Maignan, F., MacBean, N., Porcar-Castell, A., Flexas, J., Frankenberg, C., Peylin, P., Chevallier, F., Vuichard, N., Bastrikov, V., 2019. Improving Estimates of Gross Primary Productivity by Assimilating Solar-Induced Fluorescence Satellite Retrievals in a Terrestrial Biosphere Model Using a Process-Based SIF Model. Journal of Geophysical Research: Biogeosciences 124, 3281–3306. doi:10.1029/2019JG005040.

[19] Bai, J., Zhang, H., Sun, R., Li, X., Xiao, J., Wang, Y., 2022. Estimation of global GPP from GOME-2 and OCO-2 SIF by considering the dynamic variations of GPP-SIF relationship. Agricultural and Forest Meteorology 326, 109180. doi:10.1016/j.agrformet.2022.109180.

[20] Balde, H., Hmimina, G., Goulas, Y., Latouche, G., Ounis, A., Soudani, K., 2024. Data-based investigation of the effects of canopy structure and shadows on

chlorophyll fluorescence in a deciduous oak forest. Biogeosciences 21, 1259–1276. doi:10.5194/bg-21-1259-2024.

[21] Bandopadhyay, S., Rastogi, A., Juszczak, R., 2020. Review of Top-of-Canopy Sun-Induced Fluorescence (SIF) Studies from Ground, UAV, Airborne to Spaceborne Observations. Sensors 20, 1144. doi:10.3390/s20041144.

[22] Belwalkar, A., Poblete, T., Longmire, A., Hornero, A., Hernandez-Clemente, R., Zarco-Tejada, P.J., 2022. Evaluation of SIF retrievals from narrow-band and sub-nanometer airborne hyperspectral imagers flown in tandem: Modelling and validation in the context of plant phenotyping. Remote Sensing of Environment 273, 112986. doi:10.1016/j.rse.2022.112986.

[23] Bendig, J., Chang, C.Y., Wang, N., Atherton, J., Malenovský, Z., Rascher, U., 2021. Measuring Solar-Induced Fluorescence from Unmanned Aircraft Systems for Operational Use in Plant Phenotyping and Precision Farming, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 1921–1924. doi:10.1109/IGARSS47720.2021.9555157.

[24] Berger, K., Machwitz, M., Kycko, M., Kefauver, S.C., Van Wittenberghe, S., Gerhards, M., Verrelst, J., Atzberger, C., van der Tol, C., Damm, A., Rascher, U., Herrmann, I., Paz, V.S., Fahrner, S., Pieruschka, R., Prikaziuk, E., Buchaillot, M.L., Halabuk, A., Celesti, M., Koren, G., Gormus, E.T., Rossini, M., Foerster, M., Siegmann, B., Abdelbaki, A., Tagliabue, G., Hank, T., Darvishzadeh, R., Aasen, H., Garcia, M., Pôças, I., Bandopadhyay, S., Sulis, M., Tomelleri, E., Rozenstein, O., Filchev, L., Stancile, G., Schlerf, M., 2022. Multi-sensor spectral synergies for crop stress detection and monitoring in the optical domain: A review. Remote Sensing of Environment 280, 113198. doi:10.1016/j.rse.2022.113198.

[25] Berk, A., Conforti, P., Kennett, R., Perkins, T., Hawes, F., van den Bosch, J., 2014. MODTRAN® 6: A major upgrade of the MODTRAN® radiative transfer code, in: 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), pp. 1–4. doi:10.1109/WHISPERS.2014.8077573.

[26] Braham, N.A.A., Albrecht, C.M., Mairal, J., Chanussot, J., Wang, Y., Zhu, X.X., 2024. SpectralEarth: Training Hyperspectral Foundation Models at Scale. doi:10.48550/arXiv.2408.08447, arXiv:2408.08447.

[27] Brodrick, P.G., Thompson, D.R., Fahlen, J.E., Eastwood, M.L., Sarture, C.M., Lundeen, S.R., Olson-Duvall, W., Carmon, N., Green, R.O., 2021. Generalized radiative transfer emulation for imaging spectroscopy reflectance retrievals. Remote Sensing of Environment 261, 112476. doi:10.1016/j.rse.2021.112476.

[28] Bue, B.D., Thompson, D.R., Deshpande, S., Eastwood, M., Green, R.O., Natraj, V., Mullen, T., Parente, M., 2019. Neural network radiative transfer for imaging spectroscopy. Atmospheric Measurement Techniques 12, 2567–2578. doi:10.5194/amt-12-2567-2019.

[29] Buffat, J., Pato, M., Alonso, K., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H., . Emulation-based self-supervised SIF retrieval in the O2-A absorption band with HyPlant. doi:10.22541/essoar.174000855.50541566/v2.

[30] Buffat, J., Pato, M., Alonso, K., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H., 2023. Deep Learning Based Prediction of Sun-Induced Fluorescence from Hyplant Imagery, in: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, IEEE, Pasadena, CA, USA. pp. 2993–2996. doi:10.1109/IGARSS52108.2023.10282828.

[31] Buffat, J., Pato, M., Alonso, K., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H., 2024a. Retrieval of sun-induced plant fluorescence in the O2-A absorption band from DESIS imagery. doi:10.48550/arXiv.2411.08925, arXiv:2411.08925.

[32] Buffat, J., Pato, M., Alonso, K., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H., 2024b. Selected HyPlant, DESIS and OCO-3 acquisitions for the cross-comparison of sun-induced fluorescence products. doi:10.26165/JUELICH-DATA/YNYLFQ.

[33] Buffat, J., Pato, M., Alonso, K., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Siegmann, B., Rascher, U., Scharr, H., 2025. A multi-layer perceptron approach for SIF retrieval in the O2-A absorption band from hyperspectral imagery of the HyPlant airborne sensor system. Remote Sensing of Environment 318, 114596. doi:10.1016/j.rse.2024.114596.

[34] Buffat, J., Rascher, U., Rademske, P., Siegmann, B., Junker-Frohn, L.V., Emin, D., 2024c. HyData: HyPlant FLUO at-sensor radiance data packages and FLOX measurements for SIF retrieval method development from selected campaigns of the years 2018 - 2023. doi:10.26165/JUELICH-DATA/QKJKPW.

[35] Buman, B., Hueni, A., Colombo, R., Cogliati, S., Celesti, M., Julitta, T., Burkart, A., Siegmann, B., Rascher, U., Drusch, M., Damm, A., 2022. Towards consistent assessments of in situ radiometric measurements for the validation of fluorescence satellite missions. Remote Sensing of Environment 274, 112984. doi:10.1016/j.rse.2022.112984.

[36] Calderón, R., Navas-Cortés, J.A., Lucena, C., Zarco-Tejada, P.J., 2013. High-resolution airborne hyperspectral and thermal imagery for early detection of *Verticillium* wilt of olive using fluorescence, temperature and narrow-band spectral indices. Remote Sensing of Environment 139, 231–245. doi:10.1016/j.rse.2013.07.031.

[37] Calderón, R., Navas-Cortés, J.A., Zarco-Tejada, P.J., 2015. Early Detection and Quantification of Verticillium Wilt in Olive Using Hyperspectral and Thermal Imagery over Large Areas. Remote Sensing 7, 5584–5610. doi:10.3390/rs70505584.

[38] Camino, C., González-Dugo, V., Hernández, P., Sillero, J.C., Zarco-Tejada, P.J., 2018. Improved nitrogen retrievals with airborne-derived fluorescence and plant traits quantified from VNIR-SWIR hyperspectral imagery in the context of precision agriculture. International Journal of Applied Earth Observation and Geoinformation 70, 105–117. doi:10.1016/j.jag.2018.04.013.

[39] Camino, C., Gonzalez-Dugo, V., Hernandez, P., Zarco-Tejada, P.J., 2019. Radiative transfer Vcmax estimation from hyperspectral imagery and SIF retrievals

to assess photosynthetic performance in rainfed and irrigated plant phenotyping trials. Remote Sensing of Environment 231, 111186. doi:10.1016/j.rse.2019.05.005.

[40] Candiani, G., Tagliabue, G., Panigada, C., Verrelst, J., Picchi, V., Caicedo, J.P.R., Boschetti, M., 2022. Evaluation of Hybrid Models to Estimate Chlorophyll and Nitrogen Content of Maize Crops in the Framework of the Future CHIME Mission. Remote sensing 14, 1792. doi:10.3390/rs14081792.

[41] Cawse-Nicholson, K., Townsend, P.A., Schimel, D., Assiri, A.M., Blake, P.L., Buongiorno, M.F., Campbell, P., Carmon, N., Casey, K.A., Correa-Pabón, R.E., Dahlin, K.M., Dashti, H., Dennison, P.E., Dierssen, H., Erickson, A., Fisher, J.B., Frouin, R., Gatebe, C.K., Gholizadeh, H., Gierach, M., Glenn, N.F., Goodman, J.A., Griffith, D.M., Guild, L., Hakkenberg, C.R., Hochberg, E.J., Holmes, T.R.H., Hu, C., Hulley, G., Huemmrich, K.F., Kudela, R.M., Kokaly, R.F., Lee, C.M., Martin, R., Miller, C.E., Moses, W.J., Muller-Karger, F.E., Ortiz, J.D., Otis, D.B., Pahlevan, N., Painter, T.H., Pavlick, R., Poulter, B., Qi, Y., Realmuto, V.J., Roberts, D., Schaepman, M.E., Schneider, F.D., Schwandner, F.M., Serbin, S.P., Shiklomanov, A.N., Stavros, E.N., Thompson, D.R., Torres-Perez, J.L., Turpie, K.R., Tzortziou, M., Ustin, S., Yu, Q., Yusup, Y., Zhang, Q., 2021. NASA's surface biology and geology designated observable: A perspective on surface imaging algorithms. Remote Sensing of Environment 257, 112349. doi:10.1016/j.rse.2021.112349.

[42] Celesti, M., van der Tol, C., Cogliati, S., Panigada, C., Yang, P., Pinto, F., Rascher, U., Miglietta, F., Colombo, R., Rossini, M., 2018. Exploring the physiological information of Sun-induced chlorophyll fluorescence through radiative transfer model inversion. Remote Sensing of Environment 215, 97–108. doi:10.1016/j.rse.2018.05.013.

[43] Cendrero-Mateo, M., Bennertz, S., Burkart, A., Julitta, T., Cogliati, S., Scharr, H., Rademske, P., Alonso, L., Pinto, F., Rascher, U., 2018. Sun Induced Fluorescence Calibration and Validation for Field Phenotyping, in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 8248–8251. doi:10.1109/IGARSS.2018.8519174.

[44] Cendrero-Mateo, M.P., Muller, O., Albrecht, H., Burkart, A., Gatzke, S., Janssen, B., Keller, B., Körber, N., Kraska, T., Matsubara, S., Li, J., Müller-Linow, M., Pieruschka, R., Pinto, F., Rischbeck, P., Schickling, A., Steier, A., Watt, M., Schurr, U., Rascher, U., . Field Phenotyping: Concepts and Examples to Quantify Dynamic Plant Traits across Scales in the Field. Terrestrial Ecosystem Research Infrastructures .

[45] Cendrero-Mateo, M.P., Wieneke, S., Damm, A., Alonso, L., Pinto, F., Moreno, J., Guanter, L., Celesti, M., Rossini, M., Sabater, N., Cogliati, S., Julitta, T., Rascher, U., Goulas, Y., Aasen, H., Pacheco-Labrador, J., Mac Arthur, A., 2019. Sun-Induced Chlorophyll Fluorescence III: Benchmarking Retrieval Methods and Sensor Characteristics for Proximal Sensing. Remote Sensing 11, 962. doi:10.3390/rs11080962.

[46] Chang, C.Y., Guanter, L., Frankenberg, C., Köhler, P., Gu, L., Magney, T.S., Grossmann, K., Sun, Y., 2020a. Systematic Assessment of Retrieval Methods for Canopy Far-Red Solar-Induced Chlorophyll Fluorescence Using High-Frequency Automated Field Spectroscopy. Journal of Geophysical Research: Biogeosciences 125, e2019JG005533. doi:10.1029/2019JG005533.

[47] Chang, C.Y., Wen, J., Han, J., Kira, O., LeVonne, J., Melkonian, J., Riha, S.J., Skovira, J., Ng, S., Gu, L., Wood, J.D., Näthe, P., Sun, Y., 2021. Unpacking the drivers of diurnal dynamics of sun-induced chlorophyll fluorescence (SIF): Canopy structure, plant physiology, instrument configuration and retrieval methods. Remote Sensing of Environment 265, 112672. doi:10.1016/j.rse.2021.112672.

[48] Chang, C.Y., Zhou, R., Kira, O., Marri, S., Skovira, J., Gu, L., Sun, Y., 2020b. An Unmanned Aerial System (UAS) for concurrent measurements of solar-induced chlorophyll fluorescence and hyperspectral reflectance toward improving crop monitoring. Agricultural and Forest Meteorology 294, 108145. doi:10.1016/j.agrformet.2020.108145.

[49] Chapman, J.W., Thompson, D.R., Helmlinger, M.C., Bue, B.D., Green, R.O., Eastwood, M.L., Geier, S., Olson-Duvall, W., Lundeen, S.R., 2019. Spectral and Radiometric Calibration of the Next Generation Airborne Visible Infrared Spectrometer (AVIRIS-NG). Remote Sensing 11, 2129. doi:10.3390/rs11182129.

[50] Chen, R., Liu, L., Liu, X., Rascher, U., 2024. CMLR: A Mechanistic Global GPP Dataset Derived from TROPOMIS SIF Observations. Journal of Remote Sensing 4, 0127. doi:10.34133/remotesensing.0127.

[51] Chen, Y.N., Fan, K.C., Chang, Y.L., Moriyama, T., 2023. Special Issue Review: Artificial Intelligence and Machine Learning Applications in Remote Sensing. Remote Sensing 15, 569. doi:10.3390/rs15030569.

[52] Cogliati, S., Celesti, M., Cesana, I., Miglietta, F., Genesio, L., Julitta, T., Schuettemeyer, D., Drusch, M., Rascher, U., Jurado, P., Colombo, R., 2019. A Spectral Fitting Algorithm to Retrieve the Fluorescence Spectrum from Canopy Radiance. Remote Sensing 11, 1840. doi:10.3390/rs11161840.

[53] Cogliati, S., Rossini, M., Julitta, T., Meroni, M., Schickling, A., Burkart, A., Pinto, F., Rascher, U., Colombo, R., 2015a. Continuous and long-term measurements of reflectance and sun-induced chlorophyll fluorescence by using novel automated field spectroscopy systems. Remote Sensing of Environment , 12.

[54] Cogliati, S., Verhoef, W., Kraft, S., Sabater, N., Alonso, L., Vicent, J., Moreno, J., Drusch, M., Colombo, R., 2015b. Retrieval of sun-induced fluorescence using advanced spectral fitting methods. Remote Sensing of Environment 169, 344–357. doi:10.1016/j.rse.2015.08.022.

[55] Cong, Y., Khanna, S., Meng, C., Liu, P., Rozi, E., He, Y., Burke, M., Lobell, D.B., Ermon, S., 2023. SatMAE: Pre-training Transformers for Temporal and Multi-Spectral Satellite Imagery. doi:10.48550/arXiv.2207.08051, arXiv:2207.08051.

[56] Damm, A., Cogliati, S., Colombo, R., Fritsche, L., Genangeli, A., Genesio, L., Hanus, J., Peressotti, A., Rademske, P., Rascher, U., Schuettemeyer, D., Siegmann, B., Sturm, J., Miglietta, F., 2022. Response times of remote sensing measured sun-induced chlorophyll fluorescence, surface temperature and vegetation indices to evolving soil water limitation in a crop canopy. Remote Sensing of Environment 273, 112957. doi:10.1016/j.rse.2022.112957.

[57] Damm, A., Erler, A., Hillen, W., Meroni, M., Schaepman, M.E., Verhoef, W., Rascher, U., 2011. Modeling the impact of spectral sensor configurations on the FLD retrieval accuracy of sun-induced chlorophyll fluorescence. Remote Sensing of Environment 115, 1882–1892. doi:10.1016/j.rse.2011.03.011.

[58] Damm, A., Guanter, L., Laurent, V.C.E., Schaepman, M.E., Schickling, A., Rascher, U., 2014. FLD-based retrieval of sun-induced chlorophyll fluorescence from medium spectral resolution airborne spectroscopy data. Remote Sensing of Environment 147, 256–266. doi:10.1016/j.rse.2014.03.009.

[59] Damm, A., Guanter, L., Paul-Limoges, E., van der Tol, C., Hueni, A., Buchmann, N., Eugster, W., Ammann, C., Schaepman, M.E., 2015. Far-red sun-induced chlorophyll fluorescence shows ecosystem-specific relationships to gross primary production: An assessment based on observational and modeling approaches. Remote Sensing of Environment 166, 91–105. doi:10.1016/j.rse.2015.06.004.

[60] Damm, A., Haghighi, E., Paul-Limoges, E., van der Tol, C., 2021. On the seasonal relation of sun-induced chlorophyll fluorescence and transpiration in a temperate mixed forest. Agricultural and Forest Meteorology 304–305, 108386. doi:10.1016/j.agrformet.2021.108386.

[61] Damm, A., Paul-Limoges, E., Haghighi, E., Simmer, C., Morsdorf, F., Schneider, F.D., van der Tol, C., Migliavacca, M., Rascher, U., 2018. Remote sensing of plant-water relations: An overview and future perspectives. Journal of Plant Physiology 227, 3–19. doi:10.1016/j.jplph.2018.04.012.

[62] Danner, M., Berger, K., Wocher, M., Mauser, W., Hank, T., 2021. Efficient RTM-based training of machine learning regression algorithms to quantify biophysical & biochemical traits of agricultural crops. ISPRS Journal of Photogrammetry and Remote Sensing 173, 278–296. doi:10.1016/j.isprsjprs.2021.01.017.

[63] Darabi, H., Haghighi, A.T., Klöve, B., Luoto, M., 2025. Remote sensing of vegetation trends: A review of methodological choices and sources of uncertainty. Remote Sensing Applications: Society and Environment 37, 101500. doi:10.1016/j.rsase.2025.101500.

[64] Dash, J., Ogutu, B.O., 2016. Recent advances in space-borne optical remote sensing systems for monitoring global terrestrial ecosystems. Progress in Physical Geography 40, 322–351. doi:10.1177/0309133316639403.

[65] Daumard, F., Goulas, Y., Champagne, S., Fournier, A., Ounis, A., Olioso, A., Moya, I., 2012. Continuous Monitoring of Canopy Level Sun-Induced Chlorophyll Fluorescence During the Growth of a Sorghum Field. IEEE Transactions on Geoscience and Remote Sensing 50, 4292–4300. doi:10.1109/TGRS.2012.2193131.

[66] Daumard, F., Goulas, Y., Ounis, A., Pedrós, R., Moya, I., 2015. Measurement and Correction of Atmospheric Effects at Different Altitudes for Remote Sensing of Sun-Induced Fluorescence in Oxygen Absorption Bands. IEEE Transactions on Geoscience and Remote Sensing 53, 5180–5196. doi:10.1109/TGRS.2015.2418992.

[67] De Cannière, S., Herbst, M., Vereecken, H., Defourny, P., Jonard, F., 2021. Constraining water limitation of photosynthesis in a crop growth model with sun-induced chlorophyll fluorescence. Remote Sensing of Environment 267, 112722. doi:10.1016/j.rse.2021.112722.

[68] De Grave, C., Verrelst, J., Morcillo-Pallarés, P., Pipia, L., Rivera-Caicedo, J.P., Amin, E., Belda, S., Moreno, J., 2020. Quantifying vegetation biophysical variables from the Sentinel-3/FLEX tandem mission: Evaluation of the synergy of OLCI and FLORIS data sources. Remote Sensing of Environment 251, 112101. doi:10.1016/j.rse.2020.112101.

[69] Dechant, B., Ryu, Y., Badgley, G., Zeng, Y., Berry, J.A., Zhang, Y., Goulas, Y., Li, Z., Zhang, Q., Kang, M., Li, J., Moya, I., 2020. Canopy structure explains the relationship between photosynthesis and sun-induced chlorophyll fluorescence in crops. Remote Sensing of Environment 241, 111733. doi:10.1016/j.rse.2020.111733.

[70] Devlin, J., Chang, M.W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. doi:10.48550/arXiv.1810.04805, arXiv:1810.04805.

[71] Diao, W., Yu, H., Kang, K., Ling, T., Liu, D., Feng, Y., Bi, H., Ren, L., Li, X., Mao, Y., Sun, X., 2024. RingMo-Aerial: An Aerial Remote Sensing Foundation Model With A Affine Transformation Contrastive Learning. doi:10.48550/arXiv.2409.13366, arXiv:2409.13366.

[72] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N., 2021. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. doi:10.48550/arXiv.2010.11929, arXiv:2010.11929.

[73] Doughty, R., Kurosu, T.P., Parazoo, N., Köhler, P., Wang, Y., Sun, Y., Frankenberg, C., 2022. Global GOSAT, OCO-2, and OCO-3 solar-induced chlorophyll fluorescence datasets. Earth System Science Data 14, 1513–1529. doi:10.5194/essd-14-1513-2022.

[74] Dritsas, E., Trigka, M., 2025. Remote Sensing and Geospatial Analysis in the Big Data Era: A Survey. Remote Sensing 17, 550. doi:10.3390/rs17030550.

[75] Drusch, M., Moreno, J., Del Bello, U., Franco, R., Goulas, Y., Huth, A., Kraft, S., Middleton, E.M., Miglietta, F., Mohammed, G., Nedbal, L., Rascher, U., Schüttemeyer, D., Verhoef, W., 2017. The FLuorescence EXplorer Mission Concept—ESA's Earth Explorer 8. IEEE Transactions on Geoscience and Remote Sensing 55, 1273–1284. doi:10.1109/TGRS.2016.2621820.

[76] Du, K., Huang, J., Wang, W., Zeng, Y., Li, X., Zhao, F., 2024. Monitoring Low-Temperature Stress in Winter Wheat Using TROPOMI Solar-Induced Chlorophyll Fluorescence. IEEE Transactions on Geoscience and Remote Sensing 62, 1–11. doi:10.1109/TGRS.2024.3351141.

[77] Du, K., Jing, X., Zeng, Y., Ye, Q., Li, B., Huang, J., 2023a. An Improved Approach to Monitoring Wheat Stripe Rust with Sun-Induced Chlorophyll Fluorescence. Remote Sensing 15, 693. doi:10.3390/rs15030693.

[78] Du, S., Liu, L., Liu, X., Chen, J., 2021. First Investigation of the Relationship Between Solar-Induced Chlorophyll Fluorescence Observed by TanSat and Gross Primary Productivity. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 11892–11902. doi:10.1109/JSTARS.2021.3128355.

[79] Du, S., Liu, L., Liu, X., Zhang, X., Zhang, X., Bi, Y., Zhang, L., 2018. Retrieval of global terrestrial solar-induced chlorophyll fluorescence from TanSat satellite. Science Bulletin 63, 1502–1512. doi:10.1016/j.scib.2018.10.003.

[80] Du, S., Liu, X., Chen, J., Duan, W., Liu, L., 2023b. Addressing validation challenges for TROPOMI solar-induced chlorophyll fluorescence products using tower-based measurements and an NIRv-scaled approach. Remote Sensing of Environment 290, 113547. doi:10.1016/j.rse.2023.113547.

[81] Du, S., Liu, X., Duan, W., Liu, L., 2023c. Validation of solar-induced chlorophyll fluorescence products derived from OCO-2/3 observations using tower-based in situ measurements. Remote Sensing Letters 14, 713–721. doi:10.1080/2150704X.2023.2234551.

[82] Duveiller, G., Cescatti, A., 2016. Spatially downscaling sun-induced chlorophyll fluorescence leads to an improved temporal correlation with gross primary productivity. Remote Sensing of Environment 182, 72–89. doi:10.1016/j.rse.2016.04.027.

[83] Duveiller, G., Filipponi, F., Walther, S., Köhler, P., Frankenberg, C., Guanter, L., Cescatti, A., 2020. A spatially downscaled sun-induced fluorescence global product for enhanced monitoring of vegetation productivity. Earth System Science Data 12, 1101–1116. doi:10.5194/essd-12-1101-2020.

[84] Ecosystem, C.D.S., . Sentinel-2 | Copernicus Data Space Ecosystem. https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data/sentinel-2.

[85] Eldering, A., Taylor, T.E., O'Dell, C.W., Pavlick, R., 2019. The OCO-3 mission: Measurement objectives and expected performance based on 1 year of simulated data. Atmospheric Measurement Techniques 12, 2341–2370. doi:10.5194/amt-12-2341-2019.

[86] Emde, C., Buras-Schnell, R., Kylling, A., Mayer, B., Gasteiger, J., Hamann, U., Kylling, J., Richter, B., Pause, C., Dowling, T., Bugliaro, L., 2016. The libRadtran software package for radiative transfer calculations (version 2.0.1). Geoscientific Model Development 9, 1647–1672. doi:10.5194/gmd-9-1647-2016.

[87] Fletcher, K., 2015. Report for Mission Selection. Number 1330/1 in ESA SP, ESA communications production, Noordwijk.

[88] Frankenberg, C., Berry, J., 2018. Solar Induced Chlorophyll Fluorescence: Origins, Relation to Photosynthesis and Retrieval, in: Comprehensive Remote Sensing. Elsevier, pp. 143–162. doi:10.1016/B978-0-12-409548-9.10632-3.

[89] Frankenberg, C., Butz, A., Toon, G.C., 2011a. Disentangling chlorophyll fluorescence from atmospheric scattering effects in O2 A-band spectra of reflected sun-light. Geophysical Research Letters 38. doi:10.1029/2010GL045896.

[90] Frankenberg, C., Fisher, J.B., Worden, J., Badgley, G., Saatchi, S.S., Lee, J.E., Toon, G.C., Butz, A., Jung, M., Kuze, A., Yokota, T., 2011b. New global observations of the terrestrial carbon cycle from GOSAT: Patterns of plant fluorescence with gross primary productivity. Geophysical Research Letters 38. doi:10.1029/2011GL048738.

[91] Frankenberg, C., Köhler, P., Magney, T.S., Geier, S., Lawson, P., Schwochert, M., McDuffie, J., Drewry, D.T., Pavlick, R., Kuhnert, A., 2018. The Chlorophyll Fluorescence Imaging Spectrometer (CFIS), mapping far red fluorescence from aircraft. Remote Sensing of Environment 217, 523–536. doi:10.1016/j.rse.2018.08.032.

[92] Friedlingstein, P., Jones, M.W., O'Sullivan, M., Andrew, R.M., Hauck, J., Peters, G.P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D.C.E., Canadell, J.G., Ciais, P., Jackson, R.B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L.P., Currie, K.I., Feely, R.A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D.S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R.A., Hurtt, G., Ilyina, T., Jain, A.K., Joetzjer, E., Kaplan, J.O., Kato, E., Klein Goldewijk, K., Korsbakken, J.I., Landschützer, P., Lauvset, S.K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P.C., Melton, J.R., Metzl, N., Munro, D.R., Nabel, J.E.M.S., Nakaoka, S.I., Neill, C., Omar, A.M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P.P., Tian, H., Tilbrook, B., Tubiello, F.N., van der Werf, G.R., Wiltshire, A.J., Zaehle, S., 2019. Global Carbon Budget 2019. Earth System Science Data 11, 1783–1838. doi:10.5194/essd-11-1783-2019.

[93] Fu, P., Meacham-Hensold, K., Siebers, M.H., Bernacchi, C.J., 2021. The inverse relationship between solar-induced fluorescence yield and photosynthetic capacity: Benefits for field phenotyping. Journal of Experimental Botany 72, 1295–1306. doi:10.1093/jxb/eraa537.

[94] Fu, P., Montes, C.M., Siebers, M.H., Gomez-Casanovas, N., McGrath, J.M., Ainsworth, E.A., Bernacchi, C.J., 2022. Advances in field-based high-throughput photosynthetic phenotyping. Journal of Experimental Botany 73, 3157–3172. doi:10.1093/jxb/erac077.

[95] Fuller, A., Millard, K., Green, J.R., 2023. CROMA: Remote Sensing Representations with Contrastive Radar-Optical Masked Autoencoders. arXiv:2311.00566.

[96] Gamon, J.A., Somers, B., Malenovský, Z., Middleton, E.M., Rascher, U., Schaepman, M.E., 2019. Assessing Vegetation Function with Imaging Spectroscopy. Surveys in Geophysics 40, 489–513. doi:10.1007/s10712-019-09511-5.

[97] Gao, Y., Wang, S., Guan, K., Wolanin, A., You, L., Ju, W., Zhang, Y., 2020. The Ability of Sun-Induced Chlorophyll Fluorescence From OCO-2 and MODIS-EVI to Monitor Spatial Variations of Soybean and Maize Yields in the Midwestern USA. Remote Sensing 12, 1111. doi:10.3390/rs12071111.

[98] Gensheimer, J., Turner, A.J., Köhler, P., Frankenberg, C., Chen, J., 2022. A convolutional neural network for spatial downscaling of satellite-based solar-induced chlorophyll fluorescence (SIFnet). Biogeosciences 19, 1777–1793. doi:10.5194/bg-19-1777-2022.

[99] Gentine, P., Alemohammad, S.H., 2018. Reconstructed Solar-Induced Fluorescence: A Machine Learning Vegetation Product Based on MODIS Surface Reflectance to Reproduce GOME-2 Solar-Induced Fluorescence. Geophysical Research Letters 45, 3136–3146. doi:10.1002/2017GL076294.

[100] Genty, B., Briantais, J.M., Baker, N.R., 1989. The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. Biochimica et Biophysica Acta (BBA) - General Subjects 990, 87–92. doi:10.1016/S0304-4165(89)80016-9.

[101] Ghamisi, P., Gloaguen, R., Atkinson, P.M., Benediktsson, J.A., Rasti, B., Yokoya, N., Wang, Q., Hofle, B., Bruzzone, L., Bovolo, F., Chi, M., Anders, K., 2019. Multisource and Multitemporal Data Fusion in Remote Sensing: A Comprehensive Review of the State of the Art. IEEE Geoscience and Remote Sensing Magazine 7, 6–39. doi:10.1109/MGRS.2018.2890023.

[102] Ghamisi, P., Yokoya, N., Li, J., Liao, W., Liu, S., Plaza, J., Rasti, B., Plaza, A., 2017. Advances in Hyperspectral Image and Signal Processing: A Comprehensive Overview of the State of the Art. IEEE Geoscience and Remote Sensing Magazine 5, 37–78. doi:10.1109/MGRS.2017.2762087.

[103] Ghasemi, N., Justo, J.A., Celesti, M., Despoisse, L., Nieke, J., 2024. Onboard Processing of Hyperspectral Imagery: Deep Learning Advancements, Methodologies, Challenges, and Emerging Trends. https://arxiv.org/abs/2404.06526v1.

[104] Glenn, E.P., Huete, A.R., Nagler, P.L., Nelson, S.G., 2008. Relationship Between Remotely-sensed Vegetation Indices, Canopy Attributes and Plant Physiological Processes: What Vegetation Indices Can and Cannot Tell Us About the Landscape. Sensors 8, 2136–2160. doi:10.3390/s8042136.

[105] Goetz, A.F.H., 2009. Three decades of hyperspectral remote sensing of the Earth: A personal view. Remote Sensing of Environment 113, S5–S16. doi:10.1016/j.rse.2007.12.014.

[106] Gonzalez-Dugo, V., Hernandez, P., Solis, I., Zarco-Tejada, P.J., 2015. Using High-Resolution Hyperspectral and Thermal Airborne Imagery to Assess Physiological Condition in the Context of Wheat Phenotyping. Remote Sensing 7, 13586–13605. doi:10.3390/rs71013586.

[107] Goulas, Y., Fournier, A., Daumard, F., Champagne, S., Ounis, A., Marloie, O., Moya, I., 2017. Gross Primary Production of a Wheat Canopy Relates Stronger to Far Red Than to Red Solar-Induced Chlorophyll Fluorescence. Remote Sensing 9, 97. doi:10.3390/rs9010097.

[108] Grossmann, K., Frankenberg, C., Magney, T.S., Hurlock, S.C., Seibt, U., Stutz, J., 2018. PhotoSpec: A new instrument to measure spatially distributed red and far-red Solar-Induced Chlorophyll Fluorescence. Remote Sensing of Environment 216, 311–327. doi:10.1016/j.rse.2018.07.002.

[109] Gu, L., Han, J., Wood, J.D., Chang, C.Y.Y., Sun, Y., 2019a. Sun-induced Chl fluorescence and its importance for biophysical modeling of photosynthesis based on light reactions. New Phytologist 223, 1179–1191. doi:10.1111/nph.15796.

[110] Gu, L., Wood, J.D., Chang, C.Y.Y., Sun, Y., Riggs, J.S., 2019b. Advancing Terrestrial Ecosystem Science With a Novel Automated Measurement System for Sun-Induced Chlorophyll Fluorescence for Integration With Eddy Covariance Flux Networks. Journal of Geophysical Research: Biogeosciences 124, 127–146. doi:10.1029/2018JG004742.

[111] Guan, K., Berry, J.A., Zhang, Y., Joiner, J., Guanter, L., Badgley, G., Lobell, D.B., 2016. Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence. Global Change Biology 22, 716–726. doi:10.1111/gcb.13136.

[112] Guanter, L., Alonso, L., Gómez-Chova, L., Amorós-López, J., Vila, J., Moreno, J., 2007. Estimation of solar-induced vegetation fluorescence from space measurements. Geophysical Research Letters 34. doi:10.1029/2007GL029289.

[113] Guanter, L., Alonso, L., Gómez-Chova, L., Meroni, M., Preusker, R., Fischer, J., Moreno, J., 2010. Developments for vegetation fluorescence retrieval from spaceborne high-resolution spectrometry in the O2-A and O2-B absorption bands. Journal of Geophysical Research: Atmospheres 115. doi:10.1029/2009JD013716.

[114] Guanter, L., Frankenberg, C., Dudhia, A., Lewis, P.E., Gómez-Dans, J., Kuze, A., Suto, H., Grainger, R.G., 2012. Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements. Remote Sensing of Environment 121, 236–251. doi:10.1016/j.rse.2012.02.006.

[115] Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., Van der Linden, S., Leitão, P.J., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., Sang, B., 2015. The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. Remote Sensing 7, 8830–8857. doi:10.3390/rs70708830.

[116] Guanter, L., Rossini, M., Colombo, R., Meroni, M., Frankenberg, C., Lee, J.E., Joiner, J., 2013. Using field spectroscopy to assess the potential of statistical approaches for the retrieval of sun-induced chlorophyll fluorescence from ground and space. Remote Sensing of Environment 133, 52–61. doi:10.1016/j.rse.2013.01.017.

[117] Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J.A., Frankenberg, C., Huete, A.R., Zarco-Tejada, P., Lee, J.E., Moran, M.S., Ponce-Campos, G., Beer, C., Camps-Valls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker, J.M., Griffis, T.J., 2014. Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. Proceedings of the National Academy of Sciences 111, E1327–E1333. doi:10.1073/pnas.1320008111.

[118] Guerri, M.F., Distante, C., Spagnolo, P., Bougourzi, F., Taleb-Ahmed, A., 2024. Deep learning techniques for hyperspectral image analysis in agriculture: A review. ISPRS Open Journal of Photogrammetry and Remote Sensing 12, 100062. doi:10.1016/j.ophoto.2024.100062.

[119] Guillevic, P.C., Biard, J.C., Hulley, G.C., Privette, J.L., Hook, S.J., Olioso, A., Göttsche, F.M., Radocinski, R., Román, M.O., Yu, Y., Csiszar, I., 2014. Validation of Land Surface Temperature products derived from the Visible Infrared Imaging Radiometer Suite (VIIRS) using ground-based and heritage satellite measurements. Remote Sensing of Environment 154, 19–37. doi:10.1016/j.rse.2014.08.013.

[120] Guo, C., Liu, Z., Jin, X., Lu, X., 2024. Improved estimation of gross primary productivity (GPP) using solar-induced chlorophyll fluorescence (SIF) from photosystem II. Agricultural and Forest Meteorology 354, 110090. doi:10.1016/j.agrformet.2024.110090.

[121] Hao, D., Asrar, G.R., Zeng, Y., Yang, X., Li, X., Xiao, J., Guan, K., Wen, J., Xiao, Q., Berry, J.A., Chen, M., 2021. Potential of hotspot solar-induced chlorophyll fluorescence for better tracking terrestrial photosynthesis. Global Change Biology 27, 2144–2158. doi:10.1111/gcb.15554.

[122] He, L., Chen, J.M., Liu, J., Mo, G., Joiner, J., 2017. Angular normalization of GOME-2 Sun-induced chlorophyll fluorescence observation as a better proxy of vegetation productivity. Geophysical Research Letters 44, 5691–5699. doi:10.1002/2017GL073708.

[123] He, L., Magney, T., Dutta, D., Yin, Y., Köhler, P., Grossmann, K., Stutz, J., Dold, C., Hatfield, J., Guan, K., Peng, B., Frankenberg, C., 2020. From the Ground to Space: Using Solar-Induced Chlorophyll Fluorescence to Estimate Crop Productivity. Geophysical Research Letters 47, e2020GL087474. doi:10.1029/2020GL087474.

[124] Henderson, P., Ferrari, V., 2018. Learning to Generate and Reconstruct 3D Meshes with only 2D Supervision. arXiv:1807.09259.

[125] Hernández-Clemente, R., North, P.R.J., Hornero, A., Zarco-Tejada, P.J., 2017. Assessing the effects of forest health on sun-induced chlorophyll fluorescence using the FluorFLIGHT 3-D radiative transfer model to account for forest structure. Remote Sensing of Environment 193, 165–179. doi:10.1016/j.rse.2017.02.012.

[126] Hong, D., Zhang, B., Li, X., Li, Y., Li, C., Yao, J., Yokoya, N., Li, H., Ghamisi, P., Jia, X., Plaza, A., Gamba, P., Benediktsson, J.A., Chanussot, J., 2024. SpectralGPT: Spectral Remote Sensing Foundation Model. IEEE Transactions on Pattern Analysis and Machine Intelligence 46, 5227–5244. doi:10.1109/TPAMI.2024.3362475.

[127] Hong, Z., Hu, Y., Cui, C., Yang, X., Tao, C., Luo, W., Zhang, W., Li, L., Meng, L., 2022. An Operational Downscaling Method of Solar-Induced Chlorophyll Fluorescence (SIF) for Regional Drought Monitoring. Agriculture 12, 547. doi:10.3390/agriculture12040547.

[128] Hornero, A., North, P.R.J., Zarco-Tejada, P.J., Rascher, U., Martín, M.P., Migliavacca, M., Hernandez-Clemente, R., 2021. Assessing the contribution of understory sun-induced chlorophyll fluorescence through 3-D radiative transfer modelling and field data. Remote Sensing of Environment 253, 112195. doi:10.1016/j.rse.2020.112195.

[129] Hu, J., Liu, X., Liu, L., Guan, L., 2018. Evaluating the Performance of the SCOPE Model in Simulating Canopy Solar-Induced Chlorophyll Fluorescence. Remote Sensing 10, 250. doi:10.3390/rs10020250.

[130] Hu, M., Cheng, X., Zhang, J., Huang, H., Zhou, Y., Wang, X., Pan, Q., Guan, C., 2023. Temporal Variation in Tower-Based Solar-Induced Chlorophyll Fluorescence and Its Environmental Response in a Chinese Cork Oak Plantation. Remote Sensing 15, 3568. doi:10.3390/rs15143568.

[131] Huang, S.C., Pareek, A., Jensen, M., Lungren, M.P., Yeung, S., Chaudhari, A.S., 2023. Self-supervised learning for medical image classification: A systematic review and implementation guidelines. npj Digital Medicine 6, 74. doi:10.1038/s41746-023-00811-0.

[132] Israel, D., Grover, A., den Broeck, G.V., 2025. Enabling Autoregressive Models to Fill In Masked Tokens. doi:10.48550/arXiv.2502.06901, arXiv:2502.06901.

[133] Jantol, N., Prikaziuk, E., Celesti, M., Hernandez-Sequeira, I., Tomelleri, E., Pacheco-Labrador, J., Van Wittenberghe, S., Pla, F., Bandopadhyay, S., Koren, G., Siegmann, B., Legović, T., Kutnjak, H., Cendrero-Mateo, M.P., 2023. Using Sentinel-2-Based Metrics to Characterize the Spatial Heterogeneity of FLEX Sun-Induced Chlorophyll Fluorescence on Sub-Pixel Scale. Remote Sensing 15, 4835. doi:10.3390/rs15194835.

[134] JB Hyperspectral, . FloX system for Solar induced Chlorophyll Fluorescence monitoring - JB Hyperspectral - radiometer - SIF. https://www.jb-hyperspectral.com/products/flox/.

[135] Jeong, S., Ryu, Y., Li, X., Dechant, B., Liu, J., Kong, J., Choi, W., Fang, J., Lian, X., Gentine, P., 2024. GEOSIF: A continental-scale sub-daily reconstructed solar-induced fluorescence derived from OCO-3 and GK-2A over Eastern Asia and Oceania. Remote Sensing of Environment 311, 114284. doi:10.1016/j.rse.2024.114284.

[136] Jia, J., Zheng, X., Wang, Y., Chen, Y., Karjalainen, M., Dong, S., Lu, R., Wang, J., Hyyppä, J., 2024. The effect of artificial intelligence evolving on hyperspectral imagery with different signal-to-noise ratio, spectral and spatial resolutions. Remote Sensing of Environment 311, 114291. doi:10.1016/j.rse.2024.114291.

[137] Jia, M., Colombo, R., Rossini, M., Celesti, M., Zhu, J., Cogliati, S., Cheng, T., Tian, Y., Zhu, Y., Cao, W., Yao, X., 2021. Estimation of leaf nitrogen content and photosynthetic nitrogen use efficiency in wheat using sun-induced chlorophyll fluorescence at the leaf and canopy scales. European Journal of Agronomy 122, 126192. doi:10.1016/j.eja.2020.126192.

[138] Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A.P., Middleton, E.M., Huemmrich, K.F., Yoshida, Y., Frankenberg, C., 2013. Global monitoring of terrestrial chlorophyll fluorescence from moderate-spectral-resolution near-infrared satellite measurements: Methodology, simulations, and application to GOME-2. Atmospheric Measurement Techniques 6, 2803–2823. doi:10.5194/amt-6-2803-2013.

[139] Joiner, J., Yoshida, Y., Guanter, L., Lamsal, L., Li, C., Fasnacht, Z., Köhler, P., Frankenberg, C., Sun, Y., Parazoo, N., 2024. Noise reduction for solar-induced fluorescence retrievals using machine learning and principal component analysis: Simulations and applications to GOME-2 satellite retrievals. Artificial Intelligence for the Earth Systems -1. doi:10.1175/AIES-D-23-0085.1.

[140] Joiner, J., Yoshida, Y., Guanter, L., Middleton, E.M., 2016. New methods for the retrieval of chlorophyll red fluorescence from hyperspectral satellite instruments: Simulations andapplication to GOME-2 and SCIAMACHY. Atmospheric Measurement Techniques 9, 3939–3967. doi:10.5194/amt-9-3939-2016.

[141] Joiner, J., Yoshida, Y., Köehler, P., Campbell, P., Frankenberg, C., van der Tol, C., Yang, P., Parazoo, N., Guanter, L., Sun, Y., 2020. Systematic Orbital Geometry-Dependent Variations in Satellite Solar-Induced Fluorescence (SIF) Retrievals. Remote Sensing 12, 2346. doi:10.3390/rs12152346.

[142] Joiner, J., Yoshida, Y., Vasilkov, A.P., Middleton, E.M., Campbell, P.K.E., Yoshida, Y., Kuze, A., Corp, L.A., 2012. Filling-in of near-infrared solar lines by terrestrial fluorescence and other geophysical effects: Simulations and space-based observations from SCIAMACHY and GOSAT. Atmospheric Measurement Techniques 5, 809–829. doi:10.5194/amt-5-809-2012.

[143] Joiner, J., Yoshida, Y., Vasilkov, A.P., Yoshida, Y., Corp, L.A., Middleton, E.M., 2011. First observations of global and seasonal terrestrial chlorophyll fluorescence from space. Biogeosciences 8, 637–651. doi:10.5194/bg-8-637-2011.

[144] Julitta, T., Corp, L.A., Rossini, M., Burkart, A., Cogliati, S., Davies, N., Hom, M., Mac Arthur, A., Middleton, E.M., Rascher, U., Schickling, A., Colombo, R., 2016. Comparison of Sun-Induced Chlorophyll Fluorescence Estimates Obtained from Four Portable Field Spectroradiometers. Remote Sensing 8, 122. doi:10.3390/rs8020122.

[145] Jung, M., Reichstein, M., Margolis, H.A., Cescatti, A., Richardson, A.D., Arain, M.A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B.E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E.J., Papale, D., Sottocornola, M., Vaccari, F., Williams, C., 2011. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. Journal of Geophysical Research: Biogeosciences 116. doi:10.1029/2010JG001566.

[146] Kaplan, J., McCandlish, S., Henighan, T., Brown, T.B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., Amodei, D., 2020. Scaling Laws for Neural Language Models. doi:10.48550/arXiv.2001.08361, arXiv:2001.08361.

[147] Khosravi, N., Vountas, M., Rozanov, V.V., Bracher, A., Wolanin, A., Burrows, J.P., 2015. Retrieval of Terrestrial Plant Fluorescence Based on the In-Filling of Far-Red Fraunhofer Lines Using SCIAMACHY Observations. Frontiers in Environmental Science 3. doi:10.3389/fenvs.2015.00078.

[148] Kimm, H., Guan, K., Burroughs, C.H., Peng, B., Ainsworth, E.A., Bernacchi, C.J., Moore, C.E., Kumagai, E., Yang, X., Berry, J.A., Wu, G., 2021. Quantifying high-temperature stress on soybean canopy photosynthesis: The unique role of sun-induced chlorophyll fluorescence. Global Change Biology 27, 2403–2415. doi:10.1111/gcb.15603.

[149] Kira, O., Wen, J., Han, J., McDonald, A.J., Barrett, C.B., Ortiz-Bobea, A., Liu, Y., You, L., Mueller, N.D., Sun, Y., 2024. A scalable crop yield estimation framework based on remote sensing of solar-induced chlorophyll fluorescence (SIF). Environmental Research Letters 19, 044071. doi:10.1088/1748-9326/ad3142.

[150] Knopf, O., Castro, A., Bendig, J., Pude, R., Kleist, E., Poorter, H., Rascher, U., Muller, O., 2024. Field phenotyping of ten wheat cultivars under elevated CO2 shows seasonal differences in chlorophyll fluorescence, plant height and vegetation indices. Frontiers in Plant Science 14. doi:10.3389/fpls.2023.1304751.

[151] Köhler, P., Frankenberg, C., Magney, T.S., Guanter, L., Joiner, J., Landgraf, J., 2018. Global Retrievals of Solar-Induced Chlorophyll Fluorescence With TROPOMI: First Results and Intersensor Comparison to OCO-2. Geophysical Research Letters 45. doi:10.1029/2018GL079031.

[152] Köhler, P., Guanter, L., Frankenberg, C., 2015a. Simplified physically based retrieval of sun-induced chlorophyll fluorescence from GOSAT data. IEEE Geoscience and Remote Sensing Letters 12, 1446–1450. doi:10.1109/LGRS.2015.2407051.

[153] Köhler, P., Guanter, L., Joiner, J., 2015b. A linear method for the retrieval of sun-induced chlorophyll fluorescence from GOME-2 and SCIAMACHY data. Atmospheric Measurement Techniques 8, 2589–2608. doi:10.5194/amt-8-2589-2015.

[154] Kotchenova, S.Y., Vermote, E.F., 2007. Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part II. Homogeneous Lambertian and anisotropic surfaces. Applied Optics 46, 4455–4464. doi:10.1364/AO.46.004455.

[155] Kotchenova, S.Y., Vermote, E.F., Matarrese, R., Frank J. Klemm, Jr., 2006. Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path radiance. Applied Optics 45, 6762–6774. doi:10.1364/AO.45.006762.

[156] Krämer, J., Siegmann, B., Kraska, T., Muller, O., Rascher, U., 2021. The potential of spatial aggregation to extract remotely sensed sun-induced fluorescence (SIF) of small-sized experimental plots for applications in crop phenotyping. International Journal of Applied Earth Observation and Geoinformation 104, 102565. doi:10.1016/j.jag.2021.102565.

[157] Krizhevsky, A., Sutskever, I., Hinton, G.E., 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM 60, 84–90. doi:10.1145/3065386.

[158] Krutz, D., Müller, R., Knodt, U., Günther, B., Walter, I., Sebastian, I., Säuberlich, T., Reulke, R., Carmona, E., Eckardt, A., Venus, H., Fischer, C., Zender, B., Arloth, S., Lieder, M., Neidhardt, M., Grote, U., Schrandt, F., Gelmi, S., Wojtkowiak, A., 2019. The Instrument Design of the DLR Earth Sensing Imaging Spectrometer (DESIS). Sensors 19, 1622. doi:10.3390/s19071622.

[159] Lacoste, A., Lehmann, N., Rodriguez, P., Sherwin, E.D., Kerner, H., Lütjens, B., Irvin, J.A., Dao, D., Alemohammad, H., Drouin, A., Gunturkun, M., Huang, G., Vazquez, D., Newman, D., Bengio, Y., Ermon, S., Zhu, X.X., 2023. GEO-Bench: Toward Foundation Models for Earth Monitoring. doi:10.48550/arXiv.2306.03831, arXiv:2306.03831.

[160] Lenton, T.M., Abrams, J.F., Bartsch, A., Bathiany, S., Boulton, C.A., Buxton, J.E., Conversi, A., Cunliffe, A.M., Hebden, S., Lavergne, T., Poulter, B., Shepherd, A., Smith, T., Swingedouw, D., Winkelmann, R., Boers, N., 2024. Remotely sensing potential climate change tipping points across scales. Nature Communications 15, 343. doi:10.1038/s41467-023-44609-w.

[161] Li, W., Yang, W., Liu, T., Hou, Y., Li, Y., Liu, Z., Liu, Y., Liu, L., 2024a. Predicting gradient is better: Exploring self-supervised learning for SAR ATR with a joint-embedding predictive architecture. ISPRS Journal of Photogrammetry and Remote Sensing 218, 326–338. doi:10.1016/j.isprsjprs.2024.09.013.

[162] Li, X., Xiao, J., 2019. Mapping Photosynthesis Solely from Solar-Induced Chlorophyll Fluorescence: A Global, Fine-Resolution Dataset of Gross Primary Production Derived from OCO-2. Remote Sensing 11, 2563. doi:10.3390/rs11212563.

[163] Li, X., Xiao, J., He, B., Arain, M., Beringer, J., Desai, A., Emmel (nee Deutscher), C., Hollinger, D., Krasnova, A., Mammarella, I., Noe, S., Ortiz, P., Rey-Sanchez, A.C., Rocha, A., Varlagin, A., 2018. Solar-induced chlorophyll fluorescence is strongly correlated with terrestrial photosynthesis for a wide variety of biomes: First global analysis based on OCO-2 and flux tower observations. Global Change Biology 24. doi:10.1111/gcb.14297.

[164] Li, Z., Hou, B., Ma, S., Wu, Z., Guo, X., Ren, B., Jiao, L., 2024b. Masked Angle-Aware Autoencoder for Remote Sensing Images. doi:10.48550/arXiv.2408.01946, arXiv:2408.01946.

[165] Liangyun Liu, Yongjiang Zhang, Jihua Wang, Chunjiang Zhao, 2005. Detecting solar-induced chlorophyll fluorescence from field radiance spectra based on the Fraunhofer line principle. IEEE Transactions on Geoscience and Remote Sensing 43, 827–832. doi:10.1109/TGRS.2005.843320.

[166] Liu, L., Liu, X., Hu, J., 2015a. Effects of spectral resolution and SNR on the vegetation solar-induced fluorescence retrieval using FLD-based methods at canopy level. European Journal of Remote Sensing 48, 743–762. doi:10.5721/EuJRS20154841.

[167] Liu, L., Liu, X., Wang, Z., Zhang, B., 2016. Measurement and Analysis of Bidirectional SIF Emissions in Wheat Canopies. IEEE Transactions on Geoscience and Remote Sensing 54, 2640–2651. doi:10.1109/TGRS.2015.2504089.

[168] Liu, X., Liu, L., 2014. Assessing Band Sensitivity to Atmospheric Radiation Transfer for Space-Based Retrieval of Solar-Induced Chlorophyll Fluorescence. Remote Sensing 6, 10656–10675. doi:10.3390/rs61110656.

[169] Liu, X., Liu, L., Du, S., Qi, M., 2025. Inconsistent Diurnal Patterns of Far-Red Solar-Induced Chlorophyll Fluorescence Retrieved with Different Algorithms from Tower-Based Observations. Journal of Remote Sensing 5, 0429. doi:10.34133/remotesensing.0429.

[170] Liu, X., Liu, L., Zhang, S., Zhou, X., 2015b. New Spectral Fitting Method for Full-Spectrum Solar-Induced Chlorophyll Fluorescence Retrieval Based on Principal Components Analysis. Remote Sensing 7, 10626–10645. doi:10.3390/rs70810626.

[171] Liu, Y., Wang, S., Wang, X., Chen, B., Chen, J., Wang, J., Huang, M., Wang, Z., Ma, L., Wang, P., Amir, M., Zhu, K., 2022a. Exploring the superiority of solar-induced chlorophyll fluorescence data in predicting wheat yield using machine learning and deep learning methods. Computers and Electronics in Agriculture 192, 106612. doi:10.1016/j.compag.2021.106612.

[172] Liu, Z., He, X., Yang, P., Jiang, H., Xu, S., Zhao, H., Ren, S., Chen, M., 2023. Diurnal Pattern of Sun-Induced Chlorophyll Fluorescence as Reliable Indicators of Crop Water Stress. IEEE Geoscience and Remote Sensing Letters 20, 1–5. doi:10.1109/LGRS.2023.3300149.

[173] Liu, Z., Zhao, F., Liu, X., Yu, Q., Wang, Y., Peng, X., Cai, H., Lu, X., 2022b. Direct estimation of photosynthetic CO2 assimilation from solar-induced chlorophyll fluorescence (SIF). Remote Sensing of Environment 271, 112893. doi:10.1016/j.rse.2022.112893.

[174] Lone, Z.A., Pais, A.R., 2022. Object detection in hyperspectral images. Digital Signal Processing 131, 103752. doi:10.1016/j.dsp.2022.103752.

[175] Lu, X., Liu, Z., Zhao, F., Tang, J., 2020. Comparison of total emitted solar-induced chlorophyll fluorescence (SIF) and top-of-canopy (TOC) SIF in estimating photosynthesis. Remote Sensing of Environment 251, 112083. doi:10.1016/j.rse.2020.112083.

[176] Lu, X., Liu, Z., Zhou, Y., Liu, Y., An, S., Tang, J., 2018. Comparison of Phenology Estimated from Reflectance-Based Indices and Solar-Induced Chlorophyll Fluorescence (SIF) Observations in a Temperate Forest Using GPP-Based Phenology as the Standard. Remote Sensing 10, 932. doi:10.3390/rs10060932.

[177] Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing 152, 166–177. doi:10.1016/j.isprsjprs.2019.04.015.

[178] Ma, Y., Liu, L., Chen, R., Du, S., Liu, X., 2020. Generation of a Global Spatially Continuous TanSat Solar-Induced Chlorophyll Fluorescence Product by Considering the Impact of the Solar Radiation Intensity. Remote Sensing 12, 2167. doi:10.3390/rs12132167.

[179] Maes, W.H., Pagán, B.R., Martens, B., Gentine, P., Guanter, L., Steppe, K., Verhoest, N.E.C., Dorigo, W., Li, X., Xiao, J., Miralles, D.G., 2020. Sun-induced fluorescence closely linked to ecosystem transpiration as evidenced by satellite data and radiative transfer models. Remote Sensing of Environment 249, 112030. doi:10.1016/j.rse.2020.112030.

[180] Magney, T.S., Bowling, D.R., Logan, B.A., Grossmann, K., Stutz, J., Blanken, P.D., Burns, S.P., Cheng, R., Garcia, M.A., K¨hler, P., Lopez, S., Parazoo, N.C., Raczka, B., Schimel, D., Frankenberg, C., 2019a. Mechanistic evidence for tracking the seasonality of photosynthesis with solar-induced fluorescence. Proceedings of the National Academy of Sciences 116, 11640–11645. doi:10.1073/pnas.1900278116.

[181] Magney, T.S., Frankenberg, C., Köhler, P., North, G., Davis, T.S., Dold, C., Dutta, D., Fisher, J.B., Grossmann, K., Harrington, A., Hatfield, J., Stutz, J., Sun, Y., Porcar-Castell, A., 2019b. Disentangling Changes in the Spectral Shape of Chlorophyll Fluorescence: Implications for Remote Sensing of Photosynthesis. Journal of Geophysical Research: Biogeosciences 124, 1491–1507. doi:10.1029/2019JG005029.

[182] Maier, S.W., Günther, K.P., Stellmes, M., 2004. Sun-Induced Fluorescence: A New Tool for Precision Farming, in: Digital Imaging and Spectral Techniques: Applications to Precision Agriculture and Crop Physiology. John Wiley & Sons, Ltd. chapter 16, pp. 207–222. doi:10.2134/asaspecpub66.c16.

[183] Malenovský, Z., Mishra, K.B., Zemek, F., Rascher, U., Nedbal, L., 2009. Scientific and technical challenges in remote sensing of plant canopy reflectance and fluorescence. Journal of Experimental Botany 60, 2987–3004. doi:10.1093/jxb/erp156.

[184] Mañas, O., Lacoste, A., Giro-i-Nieto, X., Vazquez, D., Rodriguez, P., 2021. Seasonal Contrast: Unsupervised Pre-Training from Uncurated Remote Sensing Data. doi:10.48550/arXiv.2103.16607, arXiv:2103.16607.

[185] Marrs, J.K., Jones, T.S., Allen, D.W., Hutyra, L.R., 2021. Instrumentation sensitivities for tower-based solar-induced fluorescence measurements. Remote Sensing of Environment 259, 112413. doi:10.1016/j.rse.2021.112413.

[186] Mashala, M.J., Dube, T., Mudereri, B.T., Ayisi, K.K., Ramudzuli, M.R., 2023. A Systematic Review on Advancements in Remote Sensing for Assessing and Monitoring Land Use and Land Cover Changes Impacts on Surface Water Resources in Semi-Arid Tropical Environments. Remote Sensing 15, 3926. doi:10.3390/rs15163926.

[187] Matyukira, C., and Mhangara, P., 2024. Advances in vegetation mapping through remote sensing and machine learning techniques: A scientometric review. European Journal of Remote Sensing 57, 2422330. doi:10.1080/22797254.2024.2422330.

[188] Maxwell, K., Johnson, G.N., 2000. Chlorophyll fluorescence–a practical guide. Journal of Experimental Botany 51, 659–668. doi:10.1093/jxb/51.345.659.

[189] Mazzoni, M., Meroni, M., Fortunato, C., Colombo, R., Verhoef, W., 2012. Retrieval of maize canopy fluorescence and reflectance by spectral fitting in

the O2–A absorption band. Remote Sensing of Environment 124, 72–82. doi:10.1016/j.rse.2012.04.025.

[190] Mendieta, M., Han, B., Shi, X., Zhu, Y., Chen, C., 2023. Towards Geospatial Foundation Models via Continual Pretraining. doi:10.48550/arXiv.2302.04476, arXiv:2302.04476.

[191] Meroni, M., Barducci, A., Cogliati, S., Castagnoli, F., Rossini, M., Busetto, L., Migliavacca, M., Cremonese, E., Galvagno, M., Colombo, R., di Cella, U.M., 2011. The hyperspectral irradiometer, a new instrument for long-term and unattended field spectroscopy measurements. Review of Scientific Instruments 82, 043106. doi:10.1063/1.3574360.

[192] Meroni, M., Busetto, L., Colombo, R., Guanter, L., Moreno, J., Verhoef, W., 2010. Performance of Spectral Fitting Methods for vegetation fluorescence quantification. Remote Sensing of Environment 114, 363–374. doi:10.1016/j.rse.2009.09.010.

[193] Meroni, M., Rossini, M., Guanter, L., Alonso, L., Rascher, U., Colombo, R., Moreno, J., 2009. Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications. Remote Sensing of Environment 113, 2037–2051. doi:10.1016/j.rse.2009.05.003.

[194] Mohammed, G.H., Colombo, R., Middleton, E.M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Pérez-Priego, O., Damm, A., Meroni, M., Joiner, J., Cogliati, S., Verhoef, W., Malenovský, Z., Gastellu-Etchegorry, J.P., Miller, J.R., Guanter, L., Moreno, J., Moya, I., Berry, J.A., Frankenberg, C., Zarco-Tejada, P.J., 2019. Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. Remote Sensing of Environment 231, 111177. doi:10.1016/j.rse.2019.04.030.

[195] Moreno, J.F., 2021. Advances in the Retrieval and Interpretation of Solar-Induced Vegetation Chlorophyll Fluorescence Using Passive Remote Sensing Techniques, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 1915–1918. doi:10.1109/IGARSS47720.2021.9554239.

[196] Murchie, E., Lawson, T., 2013. Chlorophyll fluorescence analysis: A guide to good practice and understanding some new applications. Journal of Experimental Botany 64, 3983–3998. doi:10.1093/jxb/ert208.

[197] Naethe, P., Julitta, T., Chang, C.Y.Y., Burkart, A., Migliavacca, M., Guanter, L., Rascher, U., 2022. A precise method unaffected by atmospheric reabsorption for ground-based retrieval of red and far-red sun-induced chlorophyll fluorescence. Agricultural and Forest Meteorology 325, 109152. doi:10.1016/j.agrformet.2022.109152.

[198] Ni, Z., Liu, Z., Li, Z.L., Nerry, F., Huo, H., Sun, R., Yang, P., Zhang, W., 2016. Investigation of atmospheric effects on retrieval of sun-induced fluorescence using hyperspectral imagery. Sensors (Switzerland) doi:10.3390/s16040480.

[199] Niro, F., Goryl, P., Dransfeld, S., Boccia, V., Gascon, F., Adams, J., Themann, B., Scifoni, S., Doxani, G., 2021. European Space Agency (ESA) Calibration/Validation Strategy for Optical Land-Imaging Satellites and Pathway towards Interoperability. Remote Sensing 13, 3003. doi:10.3390/rs13153003.

[200] Norton, A.J., Rayner, P.J., Koffi, E.N., Scholze, M., Silver, J.D., Wang, Y.P., 2019. Estimating global gross primary productivity using chlorophyll fluorescence and a data assimilation system with the BETHY-SCOPE model. Biogeosciences 16, 3069–3093. doi:10.5194/bg-16-3069-2019.

[201] OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., Bello, I., Berdine, J., Bernadett-Shapiro, G., Berner, C., Bogdonoff, L., Boiko, O., Boyd, M., Brakman, A.L., Brockman, G., Brooks, T., Brundage, M., Button, K., Cai, T., Campbell, R., Cann, A., Carey, B., Carlson, C., Carmichael, R., Chan, B., Chang, C., Chantzis, F., Chen, D., Chen, S., Chen, R., Chen, J., Chen, M., Chess, B., Cho, C., Chu, C., Chung, H.W., Cummings, D., Currier, J., Dai, Y., Decareaux, C., Degry, T., Deutsch, N., Deville, D., Dhar, A., Dohan, D., Dowling, S., Dunning, S., Ecoffet, A., Eleti, A., Eloundou, T., Farhi, D., Fedus, L., Felix, N., Fishman, S.P., Forte, J., Fulford, I., Gao, L., Georges, E., Gibson, C., Goel, V., Gogineni, T., Goh, G., Gontijo-Lopes, R., Gordon, J., Grafstein, M., Gray, S., Greene, R., Gross, J., Gu, S.S., Guo, Y., Hallacy, C., Han, J., Harris, J., He, Y., Heaton, M., Heidecke, J., Hesse, C., Hickey, A., Hickey, W., Hoeschele, P., Houghton, B., Hsu, K., Hu, S., Hu, X., Huizinga, J., Jain, S., Jain, S., Jang, J., Jiang, A., Jiang, R., Jin, H., Jin, D., Jomoto, S., Jonn, B., Jun, H., Kaftan, T., Kaiser, Ł., Kamali, A., Kanitscheider, I., Keskar, N.S., Khan, T., Kilpatrick, L., Kim, J.W., Kim, C., Kim, Y., Kirchner, J.H., Kiros, J., Knight, M., Kokotajlo, D., Kondraciuk, Ł., Kondrich, A., Konstantinidis, A., Kosic, K., Krueger, G., Kuo, V., Lampe, M., Lan, I., Lee, T., Leike, J., Leung, J., Levy, D., Li, C.M., Lim, R., Lin, M., Lin, S., Litwin, M., Lopez, T., Lowe, R., Lue, P., Makanju, A., Malfacini, K., Manning, S., Markov, T., Markovski, Y., Martin, B., Mayer, K., Mayne, A., McGrew, B., McKinney, S.M., McLeavey, C., McMillan, P., McNeil, J., Medina, D., Mehta, A., Menick, J., Metz, L., Mishchenko, A., Mishkin, P., Monaco, V., Morikawa, E., Mossing, D., Mu, T., Murati, M., Murk, O., Mély, D., Nair, A., Nakano, R., Nayak, R., Neelakantan, A., Ngo, R., Noh, H., Ouyang, L., O'Keefe, C., Pachocki, J., Paino, A., Palermo, J., Pantuliano, A., Parascandolo, G., Parish, J., Parparita, E., Passos, A., Pavlov, M., Peng, A., Perelman, A., Peres, F.d.A.B., Petrov, M., Pinto, H.P.d.O., Michael, Pokorny, Pokrass, M., Pong, V.H., Powell, T., Power, A., Power, B., Proehl, E., Puri, R., Radford, A., Rae, J., Ramesh, A., Raymond, C., Real, F., Rimbach, K., Ross, C., Rotsted, B., Roussez, H., Ryder, N., Saltarelli, M., Sanders, T., Santurkar, S., Sastry, G., Schmidt, H., Schnurr, D., Schulman, J., Selsam, D., Sheppard, K., Sherbakov, T., Shieh, J., Shoker, S., Shyam, P., Sidor, S., Sigler, E., Simens, M., Sitkin, J., Slama, K., Sohl, I., Sokolowsky, B., Song, Y., Staudacher, N., Such, F.P., Summers, N., Sutskever, I., Tang, J., Tezak, N., Thompson, M.B., Tillet, P., Tootoonchian, A., Tseng, E., Tuggle, P., Turley, N., Tworek, J., Uribe, J.F.C., Vallone, A., Vijayvergiya, A., Voss, C., Wainwright, C., Wang, J.J., Wang, A., Wang, B., Ward, J., Wei, J., Weinmann, C.J., Welihinda, A., Welinder, P., Weng, J., Weng, L., Wiethoff, M., Willner, D., Winter, C., Wolrich, S., Wong, H., Workman, L., Wu, S., Wu, J., Wu, M., Xiao, K., Xu, T., Yoo, S., Yu, K., Yuan, Q., Zaremba, W., Zellers, R., Zhang, C., Zhang, M., Zhao, S., Zheng, T., Zhuang, J., Zhuk, W., Zoph, B., 2024. GPT-4 Technical Report. doi:10.48550/arXiv.2303.08774, arXiv:2303.08774.

[202] Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., Assran, M., Ballas, N., Galuba,

W., Howes, R., Huang, P.Y., Li, S.W., Misra, I., Rabbat, M., Sharma, V., Synnaeve, G., Xu, H., Jegou, H., Mairal, J., Labatut, P., Joulin, A., Bojanowski, P., 2024. DINOv2: Learning Robust Visual Features without Supervision. doi:10.48550/arXiv.2304.07193, arXiv:2304.07193.

[203] Pacheco-Labrador, Hueni, Mihai, Sakowska, Julitta, Kuusk, Sporea, Alonso, Burkart, Cendrero-Mateo, Aasen, Goulas, Mac Arthur, 2019a. Sun-Induced Chlorophyll Fluorescence I: Instrumental Considerations for Proximal Spectroradiometers. Remote Sensing 11, 960. doi:10.3390/rs11080960.

[204] Pacheco-Labrador, J., Perez-Priego, O., El-Madany, T.S., Julitta, T., Rossini, M., Guan, J., Moreno, G., Carvalhais, N., Martín, M.P., Gonzalez-Cascon, R., Kolle, O., Reischtein, M., van der Tol, C., Carrara, A., Martini, D., Hammer, T.W., Moossen, H., Migliavacca, M., 2019b. Multiple-constraint inversion of SCOPE. Evaluating the potential of GPP and SIF for the retrieval of plant functional traits. Remote Sensing of Environment 234, 111362. doi:10.1016/j.rse.2019.111362.

[205] Pagán, B.R., Maes, W.H., Gentine, P., Martens, B., Miralles, D.G., 2019. Exploring the Potential of Satellite Solar-Induced Fluorescence to Constrain Global Transpiration Estimates. Remote Sensing 11, 413. doi:10.3390/rs11040413.

[206] Paoletti, M.E., Haut, J.M., Plaza, J., Plaza, A., 2019. Deep learning classifiers for hyperspectral imaging: A review. ISPRS Journal of Photogrammetry and Remote Sensing 158, 279–317. doi:10.1016/j.isprsjprs.2019.09.006.

[207] Papageorgiou, G.C., Tsimilli-Michael, M., Stamatakis, K., 2007. The fast and slow kinetics of chlorophyll a fluorescence induction in plants, algae and cyanobacteria: A viewpoint. Photosynthesis Research 94, 275–290. doi:10.1007/s11120-007-9193-x.

[208] Parazoo, N.C., Bowman, K., Fisher, J.B., Frankenberg, C., Jones, D.B.A., Cescatti, A., Pérez-Priego, Ó., Wohlfahrt, G., Montagnani, L., 2014. Terrestrial gross primary production inferred from satellite fluorescence and vegetation models. Global Change Biology 20, 3103–3121. doi:10.1111/gcb.12652.

[209] Parazoo, N.C., Frankenberg, C., Köhler, P., Joiner, J., Yoshida, Y., Magney, T., Sun, Y., Yadav, V., 2019. Towards a Harmonized Long-Term Spaceborne Record of Far-Red Solar-Induced Fluorescence. Journal of Geophysical Research: Biogeosciences 124, 2518–2539. doi:10.1029/2019JG005289.

[210] Pascual-Venteo, A.B., Portalés, E., Berger, K., Tagliabue, G., Garcia, J.L., Pérez-Suay, A., Rivera-Caicedo, J.P., Verrelst, J., 2022. Prototyping Crop Traits Retrieval Models for CHIME: Dimensionality Reduction Strategies Applied to PRISMA Data. Remote sensing 14, 2448. doi:10.3390/rs14102448.

[211] Pato, M., Alonso, K., Auer, S., Buffat, J., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H., 2023. Fast Machine Learning Simulator of At-Sensor Radiances for Solar-Induced Fluorescence Retrieval with DESIS and Hyplant, in: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, pp. 7563–7566. doi:10.1109/IGARSS52108.2023.10281579.

[212] Pato, M., Alonso, K., Buffat, J., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H., 2025. Simulation Framework for Solar-Induced Fluorescence Retrieval and Application to Desis and Hyplant. doi:10.2139/ssrn.5148562.

[213] Pato, M., Buffat, J., Alonso, K., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H., 2024. Physics-based Machine Learning Emulator of At-sensor Radiances for Solar-induced Fluorescence Retrieval in the O_2-A Absorption Band. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing , 1–10doi:10.1109/JSTARS.2024.3457231.

[214] Paul-Limoges, E., Damm, A., Hueni, A., Liebisch, F., Eugster, W., Schaepman, M.E., Buchmann, N., 2018. Effect of environmental conditions on sun-induced fluorescence in a mixed forest and a cropland. Remote Sensing of Environment 219, 310–323. doi:10.1016/j.rse.2018.10.018.

[215] Paynter, I., Cook, B., Corp, L., Nagol, J., McCorkel, J., 2020. Characterization of FIREFLY, an Imaging Spectrometer Designed for Remote Sensing of Solar Induced Fluorescence. Sensors 20, 4682. doi:10.3390/s20174682.

[216] Peng, B., Guan, K., Zhou, W., Jiang, C., Frankenberg, C., Sun, Y., He, L., Köhler, P., 2020. Assessing the benefit of satellite-based Solar-Induced Chlorophyll Fluorescence in crop yield prediction. International Journal of Applied Earth Observation and Geoinformation 90, 102126. doi:10.1016/j.jag.2020.102126.

[217] Pickering, M., Cescatti, A., Duveiller, G., 2022. Sun-induced fluorescence as a proxy for primary productivity across vegetation types and climates. Biogeosciences 19, 4833–4864. doi:10.5194/bg-19-4833-2022.

[218] Pierrat, Z.A., Magney, T., Maguire, A., Brissette, L., Doughty, R., Bowling, D.R., Logan, B., Parazoo, N., Frankenberg, C., Stutz, J., . Seasonal timing of fluorescence and photosynthetic yields at needle and canopy scales in evergreen needleleaf forests. Ecology n/a, e4402. doi:10.1002/ecy.4402.

[219] Pieruschka, R., Klimov, D., Berry, J.A., Osmond, C.B., Rascher, U., Kolber, Z.S., 2012. Remote Chlorophyll Fluorescence Measurements with the Laser-Induced Fluorescence Transient Approach, in: Normanly, J. (Ed.), High-Throughput Phenotyping in Plants: Methods and Protocols. Humana Press, Totowa, NJ, pp. 51–59. doi:10.1007/978-1-61779-995-2_5.

[220] Pignatti, S., Palombo, A., Pascucci, S., Romano, F., Santini, F., Simoniello, T., Umberto, A., Vincenzo, C., Acito, N., Diani, M., Matteoli, S., Corsini, G., Casa, R., De Bonis, R., Laneve, G., Ananasso, C., 2013. The PRISMA hyperspectral mission: Science activities and opportunities for agriculture and land monitoring, in: 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS, pp. 4558–4561. doi:10.1109/IGARSS.2013.6723850.

[221] Pinto, F., Celesti, M., Acebron, K., Alberti, G., Cogliati, S., Colombo, R., Juszczak, R., Matsubara, S., Miglietta, F., Palombo, A., Panigada, C., Pignatti, S., Rossini, M., Sakowska, K., Schickling, A., Schüttemeyer, D., Stróżecki, M., Tudoroiu, M., Rascher, U., 2020. Dynamics of sun-induced chlorophyll fluorescence and reflectance to detect stress-induced variations in canopy photosynthesis. Plant, Cell & Environment 43, 1637–1654. doi:10.1111/pce.13754.

[222] Pinto, F., Müller-Linow, M., Schickling, A., Cendrero-Mateo, M.P., Ballvora, A., Rascher, U., 2017. Multiangular Observation of Canopy Sun-Induced Chlorophyll Fluorescence by Combining Imaging Spectroscopy and Stereoscopy. Remote Sensing 9, 415. doi:10.3390/rs9050415.

[223] Plascyk, J.A., Gabriel, F.C., 1975. The Fraunhofer Line Discriminator MKII-An Airborne Instrument for Precise and Standardized Ecological Luminescence Measurement. IEEE Transactions on Instrumentation and Measurement 24, 306–313. doi:10.1109/TIM.1975.4314448.

[224] Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J.C., Trianni, G., 2009. Recent advances in techniques for hyperspectral image processing. Remote Sensing of Environment 113, S110–S122. doi:10.1016/j.rse.2007.07.028.

[225] Poblete, T., Camino, C., Beck, P.S.A., Hornero, A., Kattenborn, T., Saponari, M., Boscia, D., Navas-Cortes, J.A., Zarco-Tejada, P.J., 2020. Detection of *Xylella fastidiosa* infection symptoms with airborne multispectral and thermal imagery: Assessing bandset reduction performance from hyperspectral analysis. ISPRS Journal of Photogrammetry and Remote Sensing 162, 27–40. doi:10.1016/j.isprsjprs.2020.02.010.

[226] Porcar-Castell, A., Malenovský, Z., Magney, T., Van Wittenberghe, S., Fernández-Marín, B., Maignan, F., Zhang, Y., Maseyk, K., Atherton, J., Albert, L.P., Robson, T.M., Zhao, F., Garcia-Plazaola, J.I., Ensminger, I., Rajewicz, P.A., Grebe, S., Tikkanen, M., Kellner, J.R., Ihalainen, J.A., Rascher, U., Logan, B., 2021a. Beyond APAR and NPQ: Factors Coupling and Decoupling SIF and GPP Across Scales, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 1925–1927. doi:10.1109/IGARSS47720.2021.9553153.

[227] Porcar-Castell, A., Malenovský, Z., Magney, T., Van Wittenberghe, S., Fernández-Marín, B., Maignan, F., Zhang, Y., Maseyk, K., Atherton, J., Albert, L.P., Robson, T.M., Zhao, F., Garcia-Plazaola, J.I., Ensminger, I., Rajewicz, P.A., Grebe, S., Tikkanen, M., Kellner, J.R., Ihalainen, J.A., Rascher, U., Logan, B., 2021b. Chlorophyll a fluorescence illuminates a path connecting plant molecular biology to Earth-system science. Nature Plants 7, 998–1009. doi:10.1038/s41477-021-00980-4.

[228] Porcar-Castell, A., Pfündel, E., Korhonen, J.F.J., Juurola, E., 2008. A new monitoring PAM fluorometer (MONI-PAM) to study the short- and long-term acclimation of photosystem II in field conditions. Photosynthesis Research 96, 173–179. doi:10.1007/s11120-008-9292-3.

[229] Porcar-Castell, A., Tyystjärvi, E., Atherton, J., van der Tol, C., Flexas, J., Pfündel, E.E., Moreno, J., Frankenberg, C., Berry, J.A., 2014. Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: Mechanisms and challenges. Journal of Experimental Botany 65, 4065–4095. doi:10.1093/jxb/eru191.

[230] Prexl, J., Schmitt, M., 2024. SenPa-MAE: Sensor Parameter Aware Masked Autoencoder for Multi-Satellite Self-Supervised Pretraining. doi:10.48550/arXiv.2408.11000, arXiv:2408.11000.

[231] Qian, S.E., 2021. Hyperspectral Satellites, Evolution, and Development History. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 14, 7032–7056. doi:10.1109/JSTARS.2021.3090256.

[232] Qiu, B., Xue, Y., Fisher, J.B., Guo, W., Berry, J.A., Zhang, Y., 2018. Satellite Chlorophyll Fluorescence and Soil Moisture Observations Lead to Advances in the Predictive Understanding of Global Terrestrial Coupled Carbon-Water Cycles. Global Biogeochemical Cycles 32, 360–375. doi:10.1002/2017GB005744.

[233] Qiu, R., Han, G., Ma, X., Xu, H., Shi, T., Zhang, M., 2020. A Comparison of OCO-2 SIF, MODIS GPP, and GOSIF Data from Gross Primary Production (GPP) Estimation and Seasonal Cycles in North America. Remote Sensing 12, 258. doi:10.3390/rs12020258.

[234] Qiu, R., Li, X., Han, G., Xiao, J., Ma, X., Gong, W., 2022. Monitoring drought impacts on crop productivity of the U.S. Midwest with solar-induced fluorescence: GOSIF outperforms GOME-2 SIF and MODIS NDVI, EVI, and NIRv. Agricultural and Forest Meteorology 323, 109038. doi:10.1016/j.agrformet.2022.109038.

[235] Qiu, S., Han, B., Maddix, D.C., Zhang, S., Wang, Y., Wilson, A.G., 2024. Transferring Knowledge from Large Foundation Models to Small Downstream Models. doi:10.48550/arXiv.2406.07337, arXiv:2406.07337.

[236] Quintano, C., Calvo, L., Fernández-Manso, A., Suárez-Seoane, S., Fernandes, P.M., Fernández-Guisuraga, J.M., 2023. First evaluation of fire severity retrieval from PRISMA hyperspectral data. Remote Sensing of Environment 295, 113670. doi:10.1016/j.rse.2023.113670.

[237] Raji, S.N., Subhash, N., Ravi, V., Saravanan, R., Mohanan, C.N., Nita, S., Kumar, T.M., 2015. Detection of mosaic virus disease in cassava plants by sunlight-induced fluorescence imaging: A pilot study for proximal sensing. International Journal of Remote Sensing 36, 2880–2897. doi:10.1080/01431161.2015.1049382.

[238] Rascher, U., Alonso, L., Burkart, A., Cilia, C., Cogliati, S., Colombo, R., Damm, A., Drusch, M., Guanter, L., Hanus, J., Hyvärinen, T., Julitta, T., Jussila, J., Kataja, K., Kokkalis, P., Kraft, S., Kraska, T., Matveeva, M., Moreno, J., Muller, O., Panigada, C., Pikl, M., Pinto, F., Prey, L., Pude, R., Rossini, M., Schickling, A., Schurr, U., Schüttemeyer, D., Verrelst, J., Zemek, F., 2015. Sun-induced fluorescence - a new probe of photosynthesis: First maps from the imaging spectrometer HyPlant. Global Change Biology doi:10.1111/gcb.13017.

[239] Rasti, B., Hong, D., Hang, R., Ghamisi, P., Kang, X., Chanussot, J., Benediktsson, J.A., 2020. Feature Extraction for Hyperspectral Imagery: The Evolution From Shallow to Deep: Overview and Toolbox. IEEE Geoscience and Remote Sensing Magazine 8, 60–88. doi:10.1109/MGRS.2020.2979764.

[240] Rasti, B., Ulfarsson, M.O., Sveinsson, J.R., 2016. Hyperspectral Feature Extraction Using Total Variation Component Analysis. IEEE Transactions on Geoscience and Remote Sensing 54, 6976–6985. doi:10.1109/TGRS.2016.2593463.

[241] Regaieg, O., Malenovský, Z., Siegmann, B., Buffat, J., Krämer, J., Lauret, N., Le Dantec, V., 2025. DART-based temporal and spatial retrievals of

solar-induced chlorophyll fluorescence quantum efficiency from *in-situ* and airborne crop observations. Remote Sensing of Environment 319, 114636. doi:10.1016/j.rse.2025.114636.

[242] Regaieg, O., Yin, T., Malenovský, Z., Cook, B.D., Morton, D.C., Gastellu-Etchegorry, J.P., 2021. Assessing impacts of canopy 3D structure on chlorophyll fluorescence radiance and radiative budget of deciduous forest stands using DART. Remote Sensing of Environment 265, 112673. doi:10.1016/j.rse.2021.112673.

[243] Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D.N.L., Pinto, F., Pinera-Chavez, F.J., Poland, J., Rivera-Amado, C., Saint Pierre, C., Sukumaran, S., 2020. Breeder friendly phenotyping. Plant Science 295, 110396. doi:10.1016/j.plantsci.2019.110396.

[244] Rivera-Caicedo, J.P., Verrelst, J., Muñoz-Marí, J., Camps-Valls, G., Moreno, J., 2017. Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. ISPRS Journal of Photogrammetry and Remote Sensing 132, 88–101. doi:10.1016/j.isprsjprs.2017.08.012.

[245] Roitsch, T., Cabrera-Bosquet, L., Fournier, A., Ghamkhar, K., Jiménez-Berni, J., Pinto, F., Ober, E.S., 2019. Review: New sensors and data-driven approaches—A path to next generation phenomics. Plant Science 282, 2–10. doi:10.1016/j.plantsci.2019.01.011.

[246] Rossini, M., Celesti, M., Bramati, G., Migliavacca, M., Cogliati, S., Rascher, U., Colombo, R., 2022. Evaluation of the Spatial Representativeness of In Situ SIF Observations for the Validation of Medium-Resolution Satellite SIF Products. Remote Sensing 14, 5107. doi:10.3390/rs14205107.

[247] Rossini, M., Meroni, M., Celesti, M., Cogliati, S., Julitta, T., Panigada, C., Rascher, U., Van der Tol, C., Colombo, R., 2016. Analysis of Red and Far-Red Sun-Induced Chlorophyll Fluorescence and Their Ratio in Different Canopies Based on Observed and Modeled Data. Remote Sensing 8, 412. doi:10.3390/rs8050412.

[248] Running, S.W., Nemani, R.R., Heinsch, F.A., Zhao, M., Reeves, M., Hashimoto, H., 2004. A Continuous Satellite-Derived Measure of Global Terrestrial Primary Production. BioScience 54, 547–560. doi:10.1641/0006-3568(2004)054[0547:ACSMOG]2.0.CO;2.

[249] Ryu, Y., Berry, J.A., Baldocchi, D.D., 2019. What is global photosynthesis? History, uncertainties and opportunities. Remote Sensing of Environment 223, 95–114. doi:10.1016/j.rse.2019.01.016.

[250] Sabater, N., Kolmonen, P., Van Wittenberghe, S., Arola, A., Moreno, J., 2021. Challenges in the atmospheric characterization for the retrieval of spectrally resolved fluorescence and PRI region dynamics from space. Remote Sensing of Environment 254, 112226. doi:10.1016/j.rse.2020.112226.

[251] Sabater, N., Vicent, J., Alonso, L., Cogliati, S., Verrelst, J., Moreno, J., 2017. Impact of Atmospheric Inversion Effects on Solar-Induced Chlorophyll Fluorescence: Exploitation of the Apparent Reflectance as a Quality Indicator. Remote Sensing 9, 622. doi:10.3390/rs9060622.

[252] Sabater, N., Vicent, J., Alonso, L., Verrelst, J., Middleton, E.M., Porcar-Castell, A., Moreno, J., 2018. Compensation of Oxygen Transmittance Effects for Proximal Sensing Retrieval of Canopy–Leaving Sun–Induced Chlorophyll Fluorescence. Remote Sensing 10, 1551. doi:10.3390/rs10101551.

[253] Sanders, A.F.J., Verstraeten, W.W., Kooreman, M.L., Van Leth, T.C., Beringer, J., Joiner, J., 2016. Spaceborne Sun-Induced Vegetation Fluorescence Time Series from 2007 to 2015 Evaluated with Australian Flux Tower Measurements. Remote Sensing 8, 895. doi:10.3390/rs8110895.

[254] Schaaf, C.B., Gao, F., Strahler, A.H., Lucht, W., Li, X., Tsang, T., Strugnell, N.C., Zhang, X., Jin, Y., Muller, J.P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., Doll, C., Liang, S., Privette, J.L., Roy, D., 2002. First operational BRDF, albedo nadir reflectance products from MODIS. Remote Sensing of Environment .

[255] Schaepman, M.E., Jehle, M., Hueni, A., D'Odorico, P., Damma, A., Weyermann, J., Schneider, F.D., Laurent, V., Popp, C., Seidel, F.C., Lenhard, K., Gege, P., Küchler, C., Brazile, J., Kohler, P., De Vos, L., Meuleman, K., Meynart, R., Schläpfer, D., Kneubühler, M., Itten, K.I., 2015. Advanced radiometry measurements and Earth science applicationswith the Airborne Prism Experiment (APEX). Remote Sensing of Environment 158. doi:10.1016/j.rse.2014.11.014.

[256] Scheibenreif, L., Mommert, M., Borth, D., 2023. Masked Vision Transformers for Hyperspectral Image Classification, in: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), IEEE, Vancouver, BC, Canada. pp. 2166–2176. doi:10.1109/CVPRW59228.2023.00210.

[257] Servera, J.V., Alonso, L., Martino, L., Sabater, N., Verrelst, J., Camps-Valls, G., Moreno, J., 2019. Gradient-Based Automatic Lookup Table Generator for Radiative Transfer Models. IEEE Transactions on Geoscience and Remote Sensing 57, 1040–1048. doi:10.1109/TGRS.2018.2864517.

[258] Shan, N., Ju, W., Migliavacca, M., Martini, D., Guanter, L., Chen, J., Goulas, Y., Zhang, Y., 2019. Modeling canopy conductance and transpiration from solar-induced chlorophyll fluorescence. Agricultural and Forest Meteorology 268, 189–201. doi:10.1016/j.agrformet.2019.01.031.

[259] Shekede, M.D., Gwitira, I., Chemura, A., 2024. Chapter 4 - Climate variability and agriculture, in: Dube, T., Shekede, M.D., Shoko, C., Mushore, T.D. (Eds.), Remote Sensing of Climate. Elsevier, pp. 51–83. doi:10.1016/B978-0-443-21731-9.00010-7.

[260] Shekhar, A., Buchmann, N., Gharun, M., 2022. How well do recently reconstructed solar-induced fluorescence datasets model gross primary productivity? Remote Sensing of Environment 283, 113282. doi:10.1016/j.rse.2022.113282.

[261] Shirmard, H., Farahbakhsh, E., Müller, R.D., Chandra, R., 2022. A review of machine learning in processing remote sensing data for mineral exploration. Remote Sensing of Environment 268, 112750. doi:10.1016/j.rse.2021.112750.

[262] Shwartz-Ziv, R., LeCun, Y., 2023. To Compress or Not to Compress- Self-Supervised Learning and Information Theory: A Review. doi:10.48550/arXiv.2304.09355, arXiv:2304.09355.

[263] Siegmann, B., Alonso, L., Celesti, M., Cogliati, S., Colombo, R., Damm, A., Douglas, S., Guanter, L., Hanuš, J., Kataja, K., Kraska, T., Matveeva, M., Moreno, J., Muller, O., Pikl, M., Pinto, F., Quirós Vargas, J., Rademske, P., Rodriguez-Morene, F., Sabater, N., Schickling, A., Schüttemeyer, D., Zemek, F., Rascher, U., 2019. The High-Performance Airborne Imaging Spectrometer HyPlant—From Raw Images to Top-of-Canopy Reflectance and Fluorescence Products: Introduction of an Automatized Processing Chain. Remote Sensing 11, 2760. doi:10.3390/rs11232760.

[264] Siegmann, B., Cendrero-Mateo, M.P., Cogliati, S., Damm, A., Gamon, J., Herrera, D., Jedmowski, C., Junker-Frohn, L.V., Kraska, T., Muller, O., Rademske, P., van der Tol, C., Quiros-Vargas, J., Yang, P., Rascher, U., 2021. Downscaling of far-red solar-induced chlorophyll fluorescence of different crops from canopy to leaf level using a diurnal data set acquired by the airborne imaging spectrometer HyPlant. Remote Sensing of Environment 264, 112609. doi:10.1016/j.rse.2021.112609.

[265] Sioris, C.E., Courrèges-Lacoste, G.B., Stoll, M.P., 2003. Filling in of Fraunhofer lines by plant fluorescence: Simulations for a nadir-viewing satellite-borne instrument. Journal of Geophysical Research: Atmospheres 108. doi:10.1029/2001JD001321.

[266] Sirmacek, B., Vinuesa, R., 2022. Remote sensing and AI for building climate adaptation applications. Results in Engineering 15, 100524. doi:10.1016/j.rineng.2022.100524.

[267] Sloat, L.L., Lin, M., Butler, E.E., Johnson, D., Holbrook, N.M., Huybers, P.J., Lee, J.E., Mueller, N.D., 2021. Evaluating the benefits of chlorophyll fluorescence for in-season crop productivity forecasting. Remote Sensing of Environment 260, 112478. doi:10.1016/j.rse.2021.112478.

[268] Somkuti, P., Bösch, H., Feng, L., Palmer, P.I., Parker, R.J., Quaife, T., 2020. A new space-borne perspective of crop productivity variations over the US Corn Belt. Agricultural and Forest Meteorology 281, 107826. doi:10.1016/j.agrformet.2019.107826.

[269] Song, Q., Zhu, X.G., 2024. Techniques for photosynthesis phenomics: Gas exchange, fluorescence, and reflectance spectrums. Crop and Environment 3, 147–158. doi:10.1016/j.crope.2024.05.002.

[270] Southworth, J., Smith, A.C., Safaei, M., Rahaman, M., Alruzuq, A., Tefera, B.B., Muir, C.S., Herrero, H.V., 2024. Machine learning versus deep learning in land system science: A decision-making framework for effective land classification. Frontiers in Remote Sensing 5. doi:10.3389/frsen.2024.1374862.

[271] Stavros, E.N., Chrone, J., Cawse-Nicholson, K., Freeman, A., Glenn, N.F., Guild, L., Kokaly, R., Lee, C., Luvall, J., Pavlick, R., Poulter, B., Schollaert Uz, S., Serbin, S., Thompson, D.R., Townsend, P.A., Turpie, K., Yuen, K., Thome, K., Wang, W., Zareh, S.K., Nastal, J., Bearden, D., Miller, C.E., Schimel, D.,

2023. Designing an Observing System to Study the Surface Biology and Geology (SBG) of the Earth in the 2020s. Journal of Geophysical Research: Biogeosciences 128, e2021JG006471. doi:10.1029/2021JG006471.

[272] Stewart, A., Lehmann, N., Corley, I., Wang, Y., Chang, Y.C., Ait Ali Braham, N.A., Sehgal, S., Robinson, C., Banerjee, A., 2023. SSL4EO-L: Datasets and Foundation Models for Landsat Imagery. Advances in Neural Information Processing Systems 36, 59787–59807.

[273] Stoy, P.C., El-Madany, T.S., Fisher, J.B., Gentine, P., Gerken, T., Good, S.P., Klosterhalfen, A., Liu, S., Miralles, D.G., Perez-Priego, O., Rigden, A.J., Skaggs, T.H., Wohlfahrt, G., Anderson, R.G., Coenders-Gerrits, A.M.J., Jung, M., Maes, W.H., Mammarella, I., Mauder, M., Migliavacca, M., Nelson, J.A., Poyatos, R., Reichstein, M., Scott, R.L., Wolf, S., 2019. Reviews and syntheses: Turning the challenges of partitioning ecosystem evaporation and transpiration into opportunities. Biogeosciences 16, 3747–3775. doi:10.5194/bg-16-3747-2019.

[274] Stuart, M.B., McGonigle, A.J.S., Willmott, J.R., 2019. Hyperspectral Imaging in Environmental Monitoring: A Review of Recent Developments and Technological Advances in Compact Field Deployable Systems. Sensors 19, 3071. doi:10.3390/s19143071.

[275] Sun, Y., Frankenberg, C., Jung, M., Joiner, J., Guanter, L., Köhler, P., Magney, T., 2018. Overview of Solar-Induced chlorophyll Fluorescence (SIF) from the Orbiting Carbon Observatory-2: Retrieval, cross-mission comparison, and global monitoring for GPP. Remote Sensing of Environment doi:10.1016/j.rse.2018.02.016.

[276] Sun, Y., Frankenberg, C., Wood, J.D., Schimel, D.S., Jung, M., Guanter, L., Drewry, D.T., Verma, M., Porcar-Castell, A., Griffis, T.J., Gu, L., Magney, T.S., Köhler, P., Evans, B., Yuen, K., 2017. OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. Science doi:10.1126/science.aam5747.

[277] Sun, Y., Fu, R., Dickinson, R., Joiner, J., Frankenberg, C., Gu, L., Xia, Y., Fernando, N., 2015. Drought onset mechanisms revealed by satellite solar-induced chlorophyll fluorescence: Insights from two contrasting extreme events. Journal of Geophysical Research: Biogeosciences 120, 2427–2440. doi:10.1002/2015JG003150.

[278] Sun, Y., Gu, L., Wen, J., van der Tol, C., Porcar-Castell, A., Joiner, J., Chang, C.Y., Magney, T., Wang, L., Hu, L., Rascher, U., Zarco-Tejada, P., Barrett, C.B., Lai, J., Han, J., Luo, Z., 2023a. From remotely sensed solar-induced chlorophyll fluorescence to ecosystem structure, function, and service: Part I—Harnessing theory. Global Change Biology 29, 2926–2952. doi:10.1111/gcb.16634.

[279] Sun, Y., Wen, J., Gu, L., Joiner, J., Chang, C.Y., van der Tol, C., Porcar-Castell, A., Magney, T., Wang, L., Hu, L., Rascher, U., Zarco-Tejada, P., Barrett, C.B., Lai, J., Han, J., Luo, Z., 2023b. From remotely-sensed solar-induced chlorophyll fluorescence to ecosystem structure, function, and service: Part II—Harnessing data. Global Change Biology 29, 2893–2925. doi:10.1111/gcb.16646.

[280] Tagliabue, G., Boschetti, M., Bramati, G., Candiani, G., Colombo, R., Nutini, F., Pompilio, L., Rivera-Caicedo, J.P., Rossi, M., Rossini, M., Verrelst, J., Panigada, C., 2022. Hybrid retrieval of crop traits from multi-temporal PRISMA hyperspectral imagery. ISPRS Journal of Photogrammetry and Remote Sensing 187, 362–377. doi:10.1016/j.isprsjprs.2022.03.014.

[281] Taylor, T.E., Eldering, A., Merrelli, A., Kiel, M., Somkuti, P., Cheng, C., Rosenberg, R., Fisher, B., Crisp, D., Basilio, R., Bennett, M., Cervantes, D., Chang, A., Dang, L., Frankenberg, C., Haemmerle, V.R., Keller, G.R., Kurosu, T., Laughner, J.L., Lee, R., Marchetti, Y., Nelson, R.R., O'Dell, C.W., Osterman, G., Pavlick, R., Roehl, C., Schneider, R., Spiers, G., To, C., Wells, C., Wennberg, P.O., Yelamanchili, A., Yu, S., 2020. OCO-3 early mission operations and initial (vEarly) XCO2 and SIF retrievals. Remote Sensing of Environment 251, 112032. doi:10.1016/j.rse.2020.112032.

[282] Thompson, D.R., Babu, K.N., Braverman, A.J., Eastwood, M.L., Green, R.O., Hobbs, J.M., Jewell, J.B., Kindel, B., Massie, S., Mishra, M., Mathur, A., Natraj, V., Townsend, P.A., Seidel, F.C., Turmon, M.J., 2019a. Optimal estimation of spectral surface reflectance in challenging atmospheres. Remote Sensing of Environment 232, 111258. doi:10.1016/j.rse.2019.111258.

[283] Thompson, D.R., Green, R.O., Bradley, C., Brodrick, P.G., Mahowald, N., Dor, E.B., Bennett, M., Bernas, M., Carmon, N., Chadwick, K.D., Clark, R.N., Coleman, R.W., Cox, E., Diaz, E., Eastwood, M.L., Eckert, R., Ehlmann, B.L., Ginoux, P., Ageitos, M.G., Grant, K., Guanter, L., Pearlshtien, D.H., Helmlinger, M., Herzog, H., Hoefen, T., Huang, Y., Keebler, A., Kalashnikova, O., Keymeulen, D., Kokaly, R., Klose, M., Li, L., Lundeen, S.R., Meyer, J., Middleton, E., Miller, R.L., Mouroulis, P., Oaida, B., Obiso, V., Ochoa, F., Olson-Duvall, W., Okin, G.S., Painter, T.H., Pérez García-Pando, C., Pollock, R., Realmuto, V., Shaw, L., Sullivan, P., Swayze, G., Thingvold, E., Thorpe, A.K., Vannan, S., Villarreal, C., Ung, C., Wilson, D.W., Zandbergen, S., 2024. On-orbit calibration and performance of the EMIT imaging spectrometer. Remote Sensing of Environment 303, 113986. doi:10.1016/j.rse.2023.113986.

[284] Thompson, D.R., Guanter, L., Berk, A., Gao, B.C., Richter, R., Schläpfer, D., Thome, K.J., 2019b. Retrieval of Atmospheric Parameters and Surface Reflectance from Visible and Shortwave Infrared Imaging Spectroscopy Data. Surveys in Geophysics 40, 333–360. doi:10.1007/s10712-018-9488-9.

[285] Thompson, D.R., Natraj, V., Green, R.O., Helmlinger, M.C., Gao, B.C., Eastwood, M.L., 2018. Optimal estimation for imaging spectrometer atmospheric correction. Remote Sensing of Environment 216, 355–373. doi:10.1016/j.rse.2018.07.003.

[286] Tremblay, N., Wang, Z., Cerovic, Z.G., 2012. Sensing crop nitrogen status with fluorescence indicators. A review. Agronomy for Sustainable Development 32, 451–464. doi:10.1007/s13593-011-0041-1.

[287] Trim, S.A., Buffat, J., Hueni, A., 2025. Simulation of a simultaneous traceable spectroradiometric calibration of an imaging spectrometer. Applied Optics 64, 782–796. doi:10.1364/AO.547144.

[288] Tseng, G., Cartuyvels, R., Zvonkov, I., Purohit, M., Rolnick, D., Kerner, H., 2024. Lightweight, Pre-trained Transformers for Remote Sensing Timeseries. doi:10.48550/arXiv.2304.14065, arXiv:2304.14065.

[289] U.S. Geological Survey (USGS), 2025. Landsat Missions | U.S. Geological Survey. https://www.usgs.gov/landsat-missions.

[290] Ustin, S.L., Roberts, D.A., Gamon, J.A., Asner, G.P., Green, R.O., 2004. Using Imaging Spectroscopy to Study Ecosystem Processes and Properties. BioScience 54, 523–534. doi:10.1641/0006-3568(2004)054[0523:UISTSE]2.0.CO;2.

[291] van der Tol, C., Berry, J.A., Campbell, P.K.E., Rascher, U., 2014. Models of fluorescence and photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence. Journal of Geophysical Research: Biogeosciences 119, 2312–2327. doi:10.1002/2014JG002713.

[292] Van Der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., Su, Z., 2009. An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. Biogeosciences 6, 3109–3129. doi:10.5194/bg-6-3109-2009.

[293] van Schaik, E., Kooreman, M.L., Stammes, P., Tilstra, L.G., Tuinder, O.N.E., Sanders, A.F.J., Verstraeten, W.W., Lang, R., Cacciari, A., Joiner, J., Peters, W., Boersma, K.F., 2020. Improved SIFTER v2 algorithm for long-term GOME-2A satellite retrievals of fluorescence with a correction for instrument degradation. Atmospheric Measurement Techniques 13, 4295–4315. doi:10.5194/amt-13-4295-2020.

[294] Van Wittenberghe, S., Alonso, L., Verrelst, J., Moreno, J., Samson, R., 2015. Bidirectional sun-induced chlorophyll fluorescence emission is influenced by leaf structure and light scattering properties — A bottom-up approach. Remote Sensing of Environment 158, 169–179. doi:10.1016/j.rse.2014.11.012.

[295] Vargas, J.Q., Bendig, J., Mac Arthur, A., Burkart, A., Julitta, T., Maseyk, K., Thomas, R., Siegmann, B., Rossini, M., Celesti, M., Schüttemeyer, D., Kraska, T., Muller, O., Rascher, U., 2020. Unmanned Aerial Systems (UAS)-Based Methods for Solar Induced Chlorophyll Fluorescence (SIF) Retrieval with Non-Imaging Spectrometers: State of the Art. Remote Sensing 12, 1624. doi:10.3390/rs12101624.

[296] Verhoef, W., Bach, H., 2003. Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models. Remote Sensing of Environment 87, 23–41. doi:10.1016/S0034-4257(03)00143-3.

[297] Verhoef, W., Bach, H., 2012. Simulation of Sentinel-3 images by four-stream surface–atmosphere radiative transfer modeling in the optical and thermal domains. Remote Sensing of Environment 120, 197–207. doi:10.1016/j.rse.2011.10.034.

[298] Verhoef, W., van der Tol, C., Middleton, E.M., 2018. Hyperspectral radiative transfer modeling to explore the combined retrieval of biophysical parameters and canopy fluorescence from FLEX – Sentinel-3 tandem mission multi-sensor data. Remote Sensing of Environment 204, 942–963. doi:10.1016/j.rse.2017.08.006.

[299] Verma, M., Schimel, D., Evans, B., Frankenberg, C., Beringer, J., Drewry, D.T., Magney, T., Marang, I., Hutley, L., Moore, C., Eldering, A., 2017. Effect of environmental conditions on the relationship between solar-induced fluorescence and gross primary productivity at an OzFlux grassland site. Journal of Geophysical Research: Biogeosciences 122, 716–733. doi:10.1002/2016JG003580.

[300] Verrelst, J., Rivera, J.P., van der Tol, C., Magnani, F., Mohammed, G., Moreno, J., 2015. Global sensitivity analysis of the SCOPE model: What drives simulated canopy-leaving sun-induced fluorescence? Remote Sensing of Environment 166, 8–21. doi:10.1016/j.rse.2015.06.002.

[301] Verrelst, J., Rivera-Caicedo, J.P., Reyes-Muñoz, P., Morata, M., Amin, E., Tagliabue, G., Panigada, C., Hank, T., Berger, K., 2021. Mapping landscape canopy nitrogen content from space using PRISMA data. ISPRS Journal of Photogrammetry and Remote Sensing 178, 382–395. doi:10.1016/j.isprsjprs.2021.06.017.

[302] Verrelst, J., van der Tol, C., Magnani, F., Sabater, N., Rivera, J.P., Mohammed, G., Moreno, J., 2016. Evaluating the predictive power of sun-induced chlorophyll fluorescence to estimate net photosynthesis of vegetation canopies: A SCOPE modeling study. Remote Sensing of Environment 176, 139–151. doi:10.1016/j.rse.2016.01.018.

[303] Vicent, J., Sabater, N., Tenjo, C., Acarreta, J.R., Manzano, M., Rivera, J.P., Jurado, P., Franco, R., Alonso, L., Verrelst, J., Moreno, J., 2016. FLEX End-to-End Mission Performance Simulator. IEEE Transactions on Geoscience and Remote Sensing 54, 4215–4223. doi:10.1109/TGRS.2016.2538300.

[304] Wagle, P., Zhang, Y., Jin, C., Xiao, X., 2016. Comparison of solar-induced chlorophyll fluorescence, light-use efficiency, and process-based GPP models in maize. Ecological Applications 26, 1211–1222. doi:10.1890/15-1434.

[305] Walther, S., Voigt, M., Thum, T., Gonsamo, A., Zhang, Y., Köhler, P., Jung, M., Varlagin, A., Guanter, L., 2016. Satellite chlorophyll fluorescence measurements reveal large-scale decoupling of photosynthesis and greenness dynamics in boreal evergreen forests. Global Change Biology 22, 2979–2996. doi:10.1111/gcb.13200.

[306] Wang, D., Hu, M., Jin, Y., Miao, Y., Yang, J., Xu, Y., Qin, X., Ma, J., Sun, L., Li, C., Fu, C., Chen, H., Han, C., Yokoya, N., Zhang, J., Xu, M., Liu, L., Zhang, L., Wu, C., Du, B., Tao, D., Zhang, L., 2024. HyperSIGMA: Hyperspectral Intelligence Comprehension Foundation Model. doi:10.48550/arXiv.2406.11519, arXiv:2406.11519.

[307] Wang, N., Clevers, J.G.P.W., Wieneke, S., Bartholomeus, H., Kooistra, L., 2022a. Potential of UAV-based sun-induced chlorophyll fluorescence to detect water stress in sugar beet. Agricultural and Forest Meteorology 323, 109033. doi:10.1016/j.agrformet.2022.109033.

[308] Wang, N., Siegmann, B., Rascher, U., Clevers, J.G.P.W., Muller, O., Bartholomeus, H., Bendig, J., Masiliūnas, D., Pude, R., Kooistra, L., 2022b. Comparison of a UAV- and an airborne-based system to acquire far-red sun-induced chlorophyll fluorescence measurements over structurally different crops. Agricultural and Forest Meteorology 323, 109081. doi:10.1016/j.agrformet.2022.109081.

[309] Wang, N., Suomalainen, J., Bartholomeus, H., Kooistra, L., Masiliūnas, D., Clevers, J.G.P.W., 2021. Diurnal variation of sun-induced chlorophyll fluorescence of agricultural crops observed from a point-based spectrometer on a UAV. International Journal of Applied Earth Observation and Geoinformation 96, 102276. doi:10.1016/j.jag.2020.102276.

[310] Wang, N., Yang, P., Clevers, J.G.P.W., Wieneke, S., Kooistra, L., 2023. Decoupling physiological and non-physiological responses of sugar beet to water stress from sun-induced chlorophyll fluorescence. Remote Sensing of Environment 286, 113445. doi:10.1016/j.rse.2022.113445.

[311] Wang, X., Chen, J.M., Ju, W., 2020. Photochemical reflectance index (PRI) can be used to improve the relationship between gross primary productivity (GPP) and sun-induced chlorophyll fluorescence (SIF). Remote Sensing of Environment 246, 111888. doi:10.1016/j.rse.2020.111888.

[312] Wang, Y., Suarez, L., Poblete, T., Gonzalez-Dugo, V., Ryu, D., Zarco-Tejada, P.J., 2022c. Evaluating the role of solar-induced fluorescence (SIF) and plant physiological traits for leaf nitrogen assessment in almond using airborne hyperspectral imagery. Remote Sensing of Environment 279, 113141. doi:10.1016/j.rse.2022.113141.

[313] Wen, J., Köhler, P., Duveiller, G., Parazoo, N.C., Magney, T.S., Hooker, G., Yu, L., Chang, C.Y., Sun, Y., 2020. A framework for harmonizing multiple satellite instruments to generate a long-term global high spatial-resolution solar-induced chlorophyll fluorescence (SIF). Remote Sensing of Environment 239, undefined–undefined. doi:10.1016/j.rse.2020.111644.

[314] Wieneke, S., Ahrends, H., Damm, A., Pinto, F., Stadler, A., Rossini, M., Rascher, U., 2016. Airborne based spectroscopy of red and far-red sun-induced chlorophyll fluorescence: Implications for improved estimates of gross primary productivity. Remote Sensing of Environment 184, 654–667. doi:10.1016/j.rse.2016.07.025.

[315] Wohlfahrt, G., Gerdel, K., Migliavacca, M., Rotenberg, E., Tatarinov, F., Müller, J., Hammerle, A., Julitta, T., Spielmann, F.M., Yakir, D., 2018. Sun-induced fluorescence and gross primary productivity during a heat wave. Scientific Reports 8, 14169. doi:10.1038/s41598-018-32602-z.

[316] Wu, G., Guan, K., Ainsworth, E.A., Martin, D.G., Kimm, H., Yang, X., 2024. Solar-induced chlorophyll fluorescence captures the effects of elevated ozone on canopy structure and acceleration of senescence in soybean. Journal of Experimental Botany 75, 350–363. doi:10.1093/jxb/erad356.

[317] Wu, G., Jiang, C., Kimm, H., Wang, S., Bernacchi, C., Moore, C.E., Suyker, A., Yang, X., Magney, T., Frankenberg, C., Ryu, Y., Dechant, B., Guan, K., 2022. Difference in seasonal peak timing of soybean far-red SIF and GPP explained by canopy structure and chlorophyll content. Remote Sensing of Environment 279, 113104. doi:10.1016/j.rse.2022.113104.

[318] Wu, X., Xiao, Q., Wen, J., You, D., Hueni, A., 2019. Advances in quantitative remote sensing product validation: Overview and current status. Earth-Science Reviews 196, 102875. doi:10.1016/j.earscirev.2019.102875.

[319] Xiao, A., Xuan, W., Wang, J., Huang, J., Tao, D., Lu, S., Yokoya, N., 2024. Foundation Models for Remote Sensing and Earth Observation: A Survey. doi:10.48550/arXiv.2410.16602, arXiv:2410.16602.

[320] Xiong, Z., Wang, Y., Zhang, F., Stewart, A.J., Hanna, J., Borth, D., Papoutsis, I., Saux, B.L., Camps-Valls, G., Zhu, X.X., 2024. Neural Plasticity-Inspired Multimodal Foundation Model for Earth Observation. doi:10.48550/arXiv.2403.15356, arXiv:2403.15356.

[321] Xu, Y., Zhang, L., Du, B., Zhang, L., 2022. Hyperspectral Anomaly Detection Based on Machine Learning: An Overview. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 15, 3351–3364. doi:10.1109/JSTARS.2022.3167830.

[322] Yang, J., Liu, Z., Yu, Q., Lu, X., 2024. Estimation of global transpiration from remotely sensed solar-induced chlorophyll fluorescence. Remote Sensing of Environment 303, 113998. doi:10.1016/j.rse.2024.113998.

[323] Yang, P., Van Der Tol, C., 2018. Linking canopy scattering of far-red sun-induced chlorophyll fluorescence with reflectance. Remote Sensing of Environment 209, 456–467. doi:10.1016/j.rse.2018.02.029.

[324] Yang, P., van der Tol, C., Campbell, P.K.E., Middleton, E.M., 2021. Unraveling the physical and physiological basis for the solar- induced chlorophyll fluorescence and photosynthesis relationship using continuous leaf and canopy measurements of a corn crop. Biogeosciences 18, 441–465. doi:10.5194/bg-18-441-2021.

[325] Yang, P., van der Tol, C., Verhoef, W., Damm, A., Schickling, A., Kraska, T., Muller, O., Rascher, U., 2019. Using reflectance to explain vegetation biochemical and structural effects on sun-induced chlorophyll fluorescence. Remote Sensing of Environment doi:10.1016/j.rse.2018.11.039.

[326] Yang, X., Shi, H., Stovall, A., Guan, K., Miao, G., Zhang, Y., Zhang, Y., Xiao, X., Ryu, Y., Lee, J.E., 2018. FluoSpec 2—An Automated Field Spectroscopy System to Monitor Canopy Solar-Induced Fluorescence. Sensors 18, 2063. doi:10.3390/s18072063.

[327] Yang, X., Tang, J., Mustard, J.F., Lee, J.E., Rossini, M., Joiner, J., Munger, J.W., Kornfeld, A., Richardson, A.D., 2015. Solar-induced chlorophyll fluorescence that correlates with canopy photosynthesis on diurnal and seasonal scales in a temperate deciduous forest. Geophysical Research Letters 42, 2977–2987. doi:10.1002/2015GL063201.

[328] Yao, L., Liu, Y., Yang, D., Cai, Z., Wang, J., Lin, C., Lu, N., Lyu, D., Tian, L., Wang, M., Yin, Z., Zheng, Y., Wang, S., 2022. Retrieval of solar-induced chlorophyll fluorescence (SIF) from satellite measurements: Comparison of SIF between TanSat and OCO-2. Atmospheric Measurement Techniques 15, 2125–2137. doi:10.5194/amt-15-2125-2022.

[329] Yao, L., Yang, D., Liu, Y., Wang, J., Liu, L., Du, S., Cai, Z., Lu, N., Lyu, D., Wang, M., Yin, Z., Zheng, Y., 2021. A New Global Solar-induced Chlorophyll Fluorescence (SIF) Data Product from TanSat Measurements. Advances in Atmospheric Sciences 38, 341–345. doi:10.1007/s00376-020-0204-6.

[330] Yoshida, Y., Joiner, J., Tucker, C., Berry, J., Lee, J.E., Walker, G., Reichle, R., Koster, R., Lyapustin, A., Wang, Y., 2015. The 2010 Russian drought impact on satellite measurements of solar-induced chlorophyll fluorescence: Insights from modeling and comparisons with parameters derived from satellite reflectances. Remote Sensing of Environment 166, 163–177. doi:10.1016/j.rse.2015.06.008.

[331] Yu, L., Wen, J., Chang, C.Y., Frankenberg, C., Sun, Y., 2019. High-Resolution Global Contiguous SIF of OCO-2. Geophysical Research Letters 46, 1449–1458. doi:10.1029/2018GL081109.

[332] Zarco-Tejada, P., González-Dugo, M., Fereres, E., 2016. Seasonal stability of chlorophyll fluorescence quantified from airborne hyperspectral imagery as an indicator of net photosynthesis in the context of precision agriculture. Remote Sensing of Environment 179, 89–103. doi:10.1016/j.rse.2016.03.024.

[333] Zarco-Tejada, P., González-Dugo, V., Berni, J., 2012. Fluorescence, temperature and narrow-band indices acquired from a UAV platform for water stress detection using a micro-hyperspectral imager and a thermal camera. Remote Sensing of Environment 117, 322–337. doi:10.1016/j.rse.2011.10.007.

[334] Zarco-Tejada, P.J., Suárez, L., González-Dugo, V., 2013. Spatial Resolution Effects on Chlorophyll Fluorescence Retrieval in a Heterogeneous Canopy Using Hyperspectral Imagery and Radiative Transfer Simulation. IEEE Geoscience and Remote Sensing Letters 10, 937–941. doi:10.1109/LGRS.2013.2252877.

[335] Zavafer, A., Labeeuw, L., Mancilla, C., 2020. Global Trends of Usage of Chlorophyll Fluorescence and Projections for the Next Decade. Plant Phenomics 2020. doi:10.34133/2020/6293145.

[336] Zeng, Q., Wu, X., Tang, R., Pei, J., Du, X., Pan, F., Wen, J., Xiao, Q., 2024a. Validation of TROPOMI SIF Products With Improved Geolocation Match Between In Situ and Satellite Measurements. Journal of Geophysical Research: Biogeosciences 129, e2024JG008235. doi:10.1029/2024JG008235.

[337] Zeng, Q., Wu, X., Tang, R., Wang, J., Lin, X., Wen, J., Xiao, Q., 2024b. Quantification of uncertainty caused by geometric location mismatch in the validation of TROPOMI solar-induced chlorophyll fluorescence product. International Journal of Digital Earth 17, 2330688. doi:10.1080/17538947.2024.2330688.

[338] Zeng, Y., Badgley, G., Chen, M., Li, J., Anderegg, L.D.L., Kornfeld, A., Liu, Q., Xu, B., Yang, B., Yan, K., Berry, J.A., 2020. A radiative transfer model for solar induced fluorescence using spectral invariants theory. Remote Sensing of Environment 240, 111678. doi:10.1016/j.rse.2020.111678.

[339] Zeng, Y., Chen, M., Hao, D., Damm, A., Badgley, G., Rascher, U., Johnson, J.E., Dechant, B., Siegmann, B., Ryu, Y., Qiu, H., Krieger, V., Panigada, C., Celesti, M., Miglietta, F., Yang, X., Berry, J.A., 2022. Combining near-infrared radiance of vegetation and fluorescence spectroscopy to detect effects of abiotic changes and stresses. Remote Sensing of Environment 270, 112856. doi:10.1016/j.rse.2021.112856.

[340] Zhai, X., Kolesnikov, A., Houlsby, N., Beyer, L., 2022. Scaling Vision Transformers. doi:10.48550/arXiv.2106.04560, arXiv:2106.04560.

[341] Zhang, J., Gonsamo, A., Tong, X., Xiao, J., Rogers, C.A., Qin, S., Liu, P., Yu, P., Ma, P., 2023. Solar-induced chlorophyll fluorescence captures photosynthetic phenology better than traditional vegetation indices. ISPRS Journal of Photogrammetry and Remote Sensing 203, 183–198. doi:10.1016/j.isprsjprs.2023.07.021.

[342] Zhang, L., Wang, S., Huang, C., Cen, Y., Zhai, Y., Tong, Q., 2017. Retrieval of Sun-Induced Chlorophyll Fluorescence Using Statistical Method Without Synchronous Irradiance Data. IEEE Geoscience and Remote Sensing Letters 14, 384–388. doi:10.1109/LGRS.2016.2644643.

[343] Zhang, Y., Guanter, L., Berry, J.A., Joiner, J., van der Tol, C., Huete, A., Gitelson, A., Voigt, M., Köhler, P., 2014. Estimation of vegetation photosynthetic capacity from space-based measurements of chlorophyll fluorescence for terrestrial biosphere models. Global Change Biology 20, 3727–3742. doi:10.1111/gcb.12664.

[344] Zhang, Y., Joiner, J., Alemohammad, S.H., Zhou, S., Gentine, P., 2018. A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks. Biogeosciences 15, 5779–5800. doi:10.5194/bg-15-5779-2018.

[345] Zhang, Y., Liu, X., Du, S., Qi, M., Jing, X., Liu, L., 2025. Performance of Ground-Based Solar-Induced Chlorophyll Fluorescence Retrieval Algorithms at the Water Vapor Absorption Band. Sensors 25, 689. doi:10.3390/s25030689.

[346] Zhang, Y., Migliavacca, M., Penuelas, J., Ju, W., 2021a. Advances in hyperspectral remote sensing of vegetation traits and functions. Remote Sensing of Environment 252, 112121. doi:10.1016/j.rse.2020.112121.

[347] Zhang, Y., Zhang, Q., Liu, L., Zhang, Y., Wang, S., Ju, W., Zhou, G., Zhou, L., Tang, J., Zhu, X., Wang, F., Huang, Y., Zhang, Z., Qiu, B., Zhang, X., Wang, S., Huang, C., Tang, X., Zhang, J., 2021b. ChinaSpec: A Network for Long-Term Ground-Based Measurements of Solar-Induced Fluorescence in China. Journal of Geophysical Research: Biogeosciences 126, e2020JG006042. doi:10.1029/2020JG006042.

[348] Zhang, Z., Chen, J.M., Guanter, L., He, L., Zhang, Y., 2019. From Canopy-Leaving to Total Canopy Far-Red Fluorescence Emission for Remote Sensing of Photosynthesis: First Results From TROPOMI. Geophysical Research Letters 46, 12030–12040. doi:10.1029/2019GL084832.

[349] Zhang, Z., Zhang, X., Porcar-Castell, A., Chen, J.M., Ju, W., Wu, L., Wu, Y., Zhang, Y., 2022. Sun-induced chlorophyll fluorescence is more strongly related to photosynthesis with hemispherical than nadir measurements: Evidence from field observations and model simulations. Remote Sensing of Environment 279, 113118. doi:10.1016/j.rse.2022.113118.

[350] Zhang, Z., Zhang, Y., Zhang, Q., Chen, J.M., Porcar-Castell, A., Guanter, L., Wu, Y., Zhang, X., Wang, H., Ding, D., Li, Z., 2020. Assessing bi-directional effects on the diurnal cycle of measured solar-induced chlorophyll fluorescence in crop canopies. Agricultural and Forest Meteorology 295, 108147. doi:10.1016/j.agrformet.2020.108147.

[351] Zhao, F., Guo, Y., Verhoef, W., Gu, X., Liu, L., Yang, G., 2014. A Method to Reconstruct the Solar-Induced Canopy Fluorescence Spectrum from Hyperspectral Measurements. Remote Sensing 6, 10171–10192. doi:10.3390/rs61010171.

[352] Zhao, F., Ma, W., Zhao, J., Guo, Y., Tariq, M., Li, J., 2024a. Global retrieval of the spectrum of terrestrial chlorophyll fluorescence: First results with TROPOMI. Remote Sensing of Environment 300, 113903. doi:10.1016/j.rse.2023.113903.

[353] Zhao, L., Sun, R., Zhang, J., Liu, Z., Li, S., 2024b. Matching Satellite Sun-Induced Chlorophyll Fluorescence to Flux Footprints Improves Its Relationship with Gross Primary Productivity. Remote Sensing 16, 2388. doi:10.3390/rs16132388.

[354] Zhou, K., Zhang, Q., Xiong, L., Gentine, P., 2022. Estimating evapotranspiration using remotely sensed solar-induced fluorescence measurements. Agricultural and Forest Meteorology 314, 108800. doi:10.1016/j.agrformet.2021.108800.

[355] Zhu, X., Hou, Y., Zhang, Y., Lu, X., Liu, Z., Weng, Q., 2021. Potential of Sun-Induced Chlorophyll Fluorescence for Indicating Mangrove Canopy Photosynthesis. Journal of Geophysical Research: Biogeosciences 126, e2020JG006159. doi:10.1029/2020JG006159.

[356] Zou, C., Liu, L., Du, S., Liu, X., 2022. Investigating the Potential Accuracy of Spaceborne Solar-Induced Chlorophyll Fluorescence Retrieval for 12 Capable Satellites Based on Simulation Data. IEEE Transactions on Geoscience and Remote Sensing 60, 1–13. doi:10.1109/TGRS.2022.3210185.

Appendix A

Publication I: A Multi-Layer Perceptron Approach for SIF Retrieval in the O_2-A Absorption Band from Hyperspectral Imagery of the Airborne HyPlant Sensor System

Accepted Manuscript and online published in Remote Sensing of Environment, 2025;
DOI: 10.1016/j.rse.2024.114596

Jim Buffat[a], Miguel Pato[b], Kevin Alonso[c], Stefan Auer[b], Emiliano Carmona[b], Stefan Maier[b], Rupert Müller[b], Patrick Rademske[a], Bastian Siegmann[a], Uwe Rascher[a], Hanno Scharr[d]

[a]Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Jülich, Germany
[b]Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany
[c]Starion Group c/o European Space Agency (ESA), Largo Galileo Galilei, Frascati 00044, Italy
[d]Forschungszentrum Jülich GmbH, Institute of Advanced Simulations, IAS-8: Data Analytics and Machine Learning, Jülich, Germany

Corresponding author: Jim Buffat, j.buffat@fz-juelich.de

Author contributions: **Jim Buffat**: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Miguel Pato**: Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Kevin Alonso**: Writing — review & editing, Conceptualization. **Stefan Auer**: Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Emiliano Carmona**: Writing – review & editing, Conceptualization. **Stefan Maier**: Writing – review & editing, Conceptualization. **Rupert Müller**: Conceptualization. **Patrick Rademske**: Data curation. **Bastian Siegmann**: Writing – review & editing, Data curation. **Uwe Rascher**: Writing – review & editing, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization. **Hanno Scharr**: Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

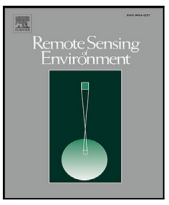
Overall contribution by Jim Buffat: 90 %

Remote Sensing of Environment 318 (2025) 114596

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse



A multi-layer perceptron approach for SIF retrieval in the O_2-A absorption band from hyperspectral imagery of the HyPlant airborne sensor system

Jim Buffat [a,d,*], Miguel Pato [b], Kevin Alonso [c], Stefan Auer [b], Emiliano Carmona [b], Stefan Maier [b], Rupert Müller [b], Patrick Rademske [a], Bastian Siegmann [a], Uwe Rascher [a], Hanno Scharr [d]

[a] Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Jülich, Germany
[b] Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany
[c] Starion Group c/o European Space Agency (ESA), Largo Galileo Galilei, Frascati 00044, Italy
[d] Forschungszentrum Jülich GmbH, Institute of Advanced Simulations, IAS-8: Data Analytics and Machine Learning, Jülich, Germany

ARTICLE INFO

Keywords:
Solar-induced fluorescence
Spectral fitting methods
Deep learning
Neural networks
Hyperspectral sensors
Remote sensing
SIF retrieval

ABSTRACT

Accurate estimation of solar-induced fluorescence (SIF) from passively sensed hyperspectral remote sensing data has been identified as fundamental in assessing the photosynthetic activity of plants for various scientific and ecological applications at different spatial scales. Different techniques to derive SIF have been developed over the last decades. In view of ESA's upcoming Earth Explorer satellite mission FLEX aiming to provide high-quality global imagery for SIF retrieval an increased interest is placed in physical approaches. We present a novel method to retrieve SIF in the O_2-A absorption band of hyperspectral imagery acquired by the HyPlant sensor system. It aims at a tight integration of physical radiative transfer principles and self-supervised neural network training. To this end, a set of spatial and spectral constraints and a specific loss formulation are adopted. In a validation study we find good agreement between our approach and established retrieval methods as well as with in-situ top-of-canopy SIF measurements. In two application studies, we additionally find evidence that the estimated SIF (i) satisfies a first-order model of diurnal SIF variation and (ii) locally adapts the estimated optical depth in topographically variable terrain.

1. Introduction

The notion that sun-induced fluorescence (SIF) is an important biophysical parameter has become prevalent over the last decades (Mohammed et al., 2019). In the last two decades, sensor and retrieval method development, on the one hand, and the push for the standardization of hyperspectral imagery in airborne and spaceborne sensing systems, on the other hand, have set the path to establish spatially contiguous SIF estimates (Grace et al., 2007; Rascher et al., 2009; Guanter et al., 2013; Ryu et al., 2019). The close causal link of the SIF signal to the internal photosynthetic machinery of plants is a valuable tool to infer plant dynamics remotely over large areas. SIF has found applications linked to establishing regional and global gross primary productivity and light-use efficiency of vegetation (Joiner et al., 2013; Cheng et al., 2013; Sun et al., 2018), small and large scale quantification of photosynthetic dynamics (Porcar-Castell et al., 2014; Rossini et al., 2015), the detection of various vegetation stress types (van der Tol et al., 2014; Verrelst et al., 2015; Damm et al., 2022), vegetative transpiration rates (Maes et al., 2020), and ecological monitoring more generally (Damm et al., 2015, 2020; Colombo et al., 2018). The retrieval of SIF from airborne hyperspectral imagery in particular allows to follow spatial and temporal patterns of photosynthetic activity at meter to submeter resolution. This provides important information for precision farming and phenotyping applications and was investigated in the context of food production (Bendig et al., 2021).

A range of hyperspectral sensors on different platforms exist for operational passive retrieval of top-of-canopy SIF (Bandopadhyay et al., 2020). In the proximal SIF retrieval, the use of field spectrometers has grown in recent years due to increased interest in passive SIF measurements as indicators of plant status in application studies. Furthermore, these ground measurements are important in the development of UAV-based, airborne and spaceborne sensors for SIF retrieval. High-quality hyperspectral ground data sets for calibration and validation of non-ground-based sensors are increasingly needed (Naethe et al., 2024) as the number of SIF products increases. In recent years, a multitude of globally distributed spaceborne SIF products have been developed from sensors that were originally designed for the characterization of the atmosphere such as GOSAT (Joiner et al., 2011), GOME (Joiner et al., 2013; Guanter et al., 2014), SCIAMACHY (Joiner et al., 2016),

* Corresponding author at: Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Jülich, Germany.
E-mail address: j.buffat@fz-juelich.de (J. Buffat).

https://doi.org/10.1016/j.rse.2024.114596
Received 13 April 2024; Received in revised form 23 November 2024; Accepted 30 December 2024
Available online 15 January 2025

OCO-2 (Sun et al., 2017), TROPOMI (Guanter et al., 2015a, 2021) and TanSAT (Yao et al., 2021). In addition, the upcoming FLEX satellite mission will make available data from the first hyperspectral sensor dedicated full-spectrum SIF retrieval on a global scale by 2026 (Drusch et al., 2017).

Investigations regarding the development, improvement and testing of operational SIF retrieval algorithms have relied strongly on data from airborne platforms equipped with sensors such as with the micro-hyperspectral imaging sensor (Zarco-Tejada et al., 2016), with the Chlorophyll Fluorescence Imaging Spectrometer (CFIS) (Frankenberg et al., 2018) and with the HyPlant spectrometer (Siegmann et al., 2019). The development of the new SIF retrieval approach presented in this work is based on hyperspectral HyPlant imagery. HyPlant data is particularly suitable to testing new retrieval methods as it has been designed as an airborne demonstrator for the FLORIS sensor of the upcoming spaceborne FLEX mission. Moreover, hundreds of HyPlant flight lines were recorded in several field campaigns since 2014. This large collection of hyperspectral data sets allowed to test different SIF retrieval methods (European Space Agency, 2017a,b, 2018, 2019; Rascher et al., 2021, 2022a,b). Many of these data sets are acquired in tandem with ground-based measurements allowing to derive high quality ground-based SIF estimates as 'ground truth' in addition to the hyperspectral flight lines (see Section 2.2 for details on the data sets and ground measurements).

Prominent algorithms for SIF retrieval on hyperspectral data are the 3FLD method (Maier et al., 2004), the Improved Fraunhofer Line Discrimination (iFLD) (Alonso et al., 2008), the Spectral Fitting Method (SFM) targeting SIF in oxygen absorption bands (Mazzoni et al., 2012; Meroni et al., 2010) as well as full-spectrum SFM (Cogliati et al., 2015b, 2018, 2019) developed in view of the ambitious accuracy requirements of the FLEX mission, the Singular Vector Decomposition (SVD) method (Guanter et al., 2012) as well as the recently developed partial least-squares (PLS) approach (Naethe et al., 2022). In this work, we examine the performance of a new spectral fitting method neural network (called SFMNN) with a novel self-supervised loss function specifically formulated for SIF retrieval from hyperspectral imagery first presented in Buffat et al. (2023) and compare it to SFM (Cogliati et al., 2015b) and iFLD (Wieneke et al., 2016) versions adapted to SIF retrieval in the O_2-A absorption band of HyPlant acquisitions. The current work restricts the model fitting to the O_2-A band which is a subset of the radiance spectrum measured by HyPlant. While restricting the spectral range decreases the amount of constraining information during training it also decreases the complexity and, thus, the required complexity of the physical model fitted to the data. Fitting the whole HyPlant spectrum with the proposed methodology for full-spectrum SIF retrieval is, however, principally possible. It can be achieved by adjusting the formulation of the physical at-sensor radiance simulation. In this contribution we have, however, opted to present a more concise retrieval problem in the O_2-A band in order to discuss our methodology.

As in other spectral fitting retrieval schemes we aim at performing a decomposition of the at-sensor radiance into reflected irradiance and the top-of-canopy fluorescence transmitted to the optical sensor. We investigate the use of a neural network to learn a feature based decomposition instead of leveraging pixelwise optimization as is implemented in other spectral fitting methods. This decomposition requires a joint estimation of the surface reflectance, the top-of-canopy fluorescence and the influence of atmospheric transmittance and scattering during data acquisition, which cannot be inferred based on physical principles alone due to incomplete knowledge of the atmospheric composition and the surface properties at acquisition time. In its most general form such a joint estimation constitutes an ill-posed inversion problem as different decompositions can result in the same observation (Frankenberg et al., 2011). We introduce constraints based on prior knowledge of the spectral forms and spatial variability of the atmospheric transfer functions, the surface reflectance and the fluorescence emission in the O_2-A absorption band to reduce the space of possible solutions. A similar set-up has been adopted in earlier SFM implementations restricted to the O_2-A and O_2-B absorption band (Mazzoni et al., 2012; Meroni et al., 2010) as well as for full-spectrum retrieval (Cogliati et al., 2015b, 2019).

Fluorescence and reflectance are spectrally smooth, but vary strongly in space depending mainly on the distribution of vegetation, the photosynthetic state of the vegetation and soil conditions. In contrast to this, atmospheric functions are spectrally highly variable but spatially smooth as they vary on spatial scales that are larger than the typical spatial footprints of airborne imagery (Anderson et al., 2003; Thompson et al., 2021) which allows the simplifying assumption of constant atmospheric conditions over extended spatial domains for the purpose of fluorescence estimation. Atmospheric transfer functions are therefore typically estimated for entire flight lines or large spatial image regions in existing SIF retrieval schemes.

To disentangle fluorescence and reflectance, iFLD and SFM formulate pixelwise functional constraints, i.e. constraints on full spectra or on a collection of spectral lines, that either allow to fit a specific at-sensor radiance model (SFM) or else to solve an equation system (iFLD). However, the assumption of constant atmospheric conditions causes performance decreases in regions with variable atmospheric transfer properties, e.g., because of varying optical path lengths due to variable topography, aerosol and water vapor concentrations, temperature and pressure profiles and clouds (Sabater et al., 2021; Guanter et al., 2010; Pato et al., 2024). Disregarding such atmospheric variability can result in misattribution of signal components in the decomposition that SIF retrieval methods perform. Changing optical path lengths results in variable absorption of oxygen and, thus, in changing O_2-A absorption depths. These changes cannot be attributed uniquely to the presence of fluorescence anymore as a result. In cases where the atmospheric transfer functions cannot be assumed sufficiently constant in the spatial dimension, localized estimates of the atmospheric transfer or an atmospheric correction of the at-sensor radiance is necessary. If retrieval methods do not rely on atmospherically corrected data, but characterize the atmosphere on-the-fly (such as the iFLD and SFM baseline methods in this contribution), they must be run repeatedly on independent localized subsets of the data where the assumption of constant topography holds approximately. This results in a computational overhead and exposes the retrieval to errors incurred due to the spatial discretization of the input data. While approaches for computationally efficient estimation of spatially variable atmospheric transfer exists (Thompson et al., 2022), such approaches have not been integrated in existing SIF retrieval methods. The introduction of atmospheric correction on the other hand exposes the SIF retrieval to errors in misparameterization of the correction algorithm, particularly in acquisition contexts where only incomplete knowledge of the atmospheric components exists (van der Tol et al., 2023). For this reason, a practical contribution of this study is that our new retrieval method allows for locally varying atmospheric conditions and, thus, is able to provide SIF estimates for topographically complex terrains. The proposed method does not rely on a separate atmospheric characterization or correction step as in SFM and iFLD since the estimation of the atmospheric transfer functions is learned as part of a joint estimation problem.

Recently published neural network based SIF retrieval schemes focus on reconstructing traditionally derived spaceborne SIF products from reflectance and PAR products and a subset of high-quality SIF estimates used as labels (Gentine and Alemohammad, 2018; Zhang et al., 2018; Chen et al., 2022). Our approach is different to these approaches as we do not use labels for training, but leverage a set of constraints to invert a physical model of the at-sensor radiance. Our approach uses similar model assumptions as SFM (Cogliati et al., 2015b) for the different parts of the at-sensor radiance model, i.e., the fluorescence, reflectance, and atmospheric functions, since the at-sensor radiance needs to be simulated efficiently during the loss minimization of SFMNN. In the case of SFMNN we additionally require

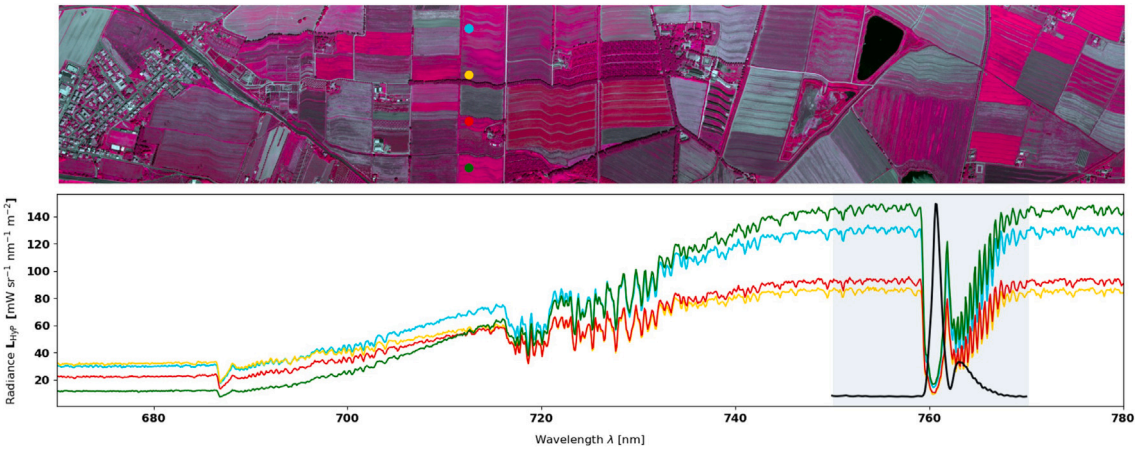


Fig. 1. Sample spectra of a WST-2019 HyPlant acquisition (false color image (FLUO NIR, red and green), upper row) over the input spectral window (lower row) $\mathcal{W}_{\text{tot}}$. The window $\mathcal{W}$ (750–770 nm) simulated by our approach is highlighted in blue. In black we show the SNR based weighting w_f described in Section 3.4.1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

a simulation model from which a gradient can be derived easily with automatic differentiation.

This work encompasses a qualitative comparison of the reconstruction residuals of SFMNN as compared to SFM and a quantitative comparison of its SIF predictions with FLOX top-of-canopy SIF measurements. Additionally, we show in two application studies that SFMNN's SIF predictions are in line with physiological and physical plausibility considerations.

In the first application study, we analyze the diurnal course of SFMNN predicted SIF. Understanding the diurnal dynamics of canopy SIF and its interplay with plant status and photosynthetic activity is an open and relevant research question (Lee et al., 2015; Kim et al., 2021; Siegmann et al., 2021; Pierrat et al., 2022). SIF as a function of daytime has been modeled in the past by van der Tol et al. (2016) and has been the subject of detailed studies more recently (Wang et al., 2021; Chang et al., 2021). As a plausibility check of SFMNN SIF predictions, we can make use of the universal feature of diurnal SIF dynamics to be strongly correlated to the local solar irradiance due to the large dependency of SIF on available photosynthetically active radiation (PAR). Given a predicted SIF time series we can formulate a first order plausibility test on whether the predicted SIF peaks around solar noon.

The SFMNN prediction of the atmospheric transfer is designed to be locally adaptive. The second application study focuses therefore on the sensitivity of the atmospheric transfer prediction to changing atmospheric and observational conditions. Due to a lack of knowledge about the atmospheric state at the time of data acquisition, we cannot compare the predicted atmospheric conditions to a measured or estimated atmospheric composition to gain insights into the physical accuracy of the predicted optical depth of the atmosphere. The only driver of the atmospheric transfer that can be quantitatively assessed by us is the optical path length of a given pixel on the ground. Therefore we test SFMNN's capacity to adapt to local changes in the optical path in HyPlant acquisitions with considerable topographic variation.

2. Data

2.1. HyPlant spectrometer

HyPlant is an assembly of two hyperspectral push-broom sensors with different spectral sampling and resolution configurations. The FLUO module covers the range 670–780 nm with a spectral sampling interval of 0.11 nm. Its resolution in terms of the full width half maximum (FWHM) is 0.27 nm (Siegmann et al., 2019). See Fig. 1 for sample spectra over the full spectral range of HyPlant. The DUAL module in contrast, provides optical data in the spectral range 370–2500 nm with 3–10 nm spectral sampling and a FWHM varying between 1.7–5.6 nm. The FLUO module was specifically designed for SIF retrieval in the O_2-A and O_2-B absorption bands when operated on airborne platforms at flight heights varying from 350 m to 1800 m. The DUAL module's purpose is to simultaneously provide optical data from which ancillary information, such as the reflectance and various related products (e.g., the photochemical reflectance index (PRI) or the Enhanced Vegetation Index (EVI) (Bandopadhyay et al., 2019)), can be derived. The current work makes use of radiometrically calibrated FLUO acquisitions in sensor geometry as input for the SIF predictor network (Buffat et al., 2024b). The input data is not spectrally calibrated. Rather, the spectral calibration is estimated as part of the SIF retrieval scheme presented here .

We train a neural network (see Fig. 2) with data from 13 HyPlant campaigns in five locations from six years listed in Table 1. The measurement campaigns used for training cover a wide range of the nominal flight heights at which HyPlant is operated (350–1800 m). The acquisitions of these campaigns also differ with respect to the variability of the ground altitude. NRS-2019, TR32-2019, GLO-2019, INN-2022 and SOP-2023 comprise acquisitions with ground altitude changes of 50–150 m, while WST-2019, SEL-2018 and the CKA campaigns were conducted over flat terrain (less than 20 m). The diversity of sun-surface-sensor geometries over this multi-year data set results in significant variability of the O_2-A absorption band depth.

The training process of our method involves a pretraining on a large and variable data set and subsequent finetuning to smaller data sets consisting of HyPlant imagery acquired at the same nominal flight height and comparable topographic variation. During pretraining the model is trained to minimize the same loss as during finetuning. Pretraining is conducted to derive a coarse base predictor which then can be finetuned to different datasets. From this base predictor data set specific models can be derived with fewer training steps than a full training process would require. We expect the variability of the fitted atmospheric transfer functions during finetuning to be smaller than in the full data set such that finetuning also aims at improving the base predictors performance for the specific atmospheric transfer conditions prevalent in different data sets.

For the pretraining we compiled a data set $\mathcal{D}_{\text{mix}}$ consisting of acquisitions with large differences in nominal flight height and topographic conditions (see Table 2). The variability of this data set ensured the generalization capacity of the coarse model to a wide variety of observation conditions during finetuning.

For finetuning, we created the data sets $\mathcal{D}_{350}$, $\mathcal{D}_{600}$ and $\mathcal{D}_{1500}$ for individual nominal flight height by selecting a subset of the available acquisitions of the measurement campaigns with small topographic

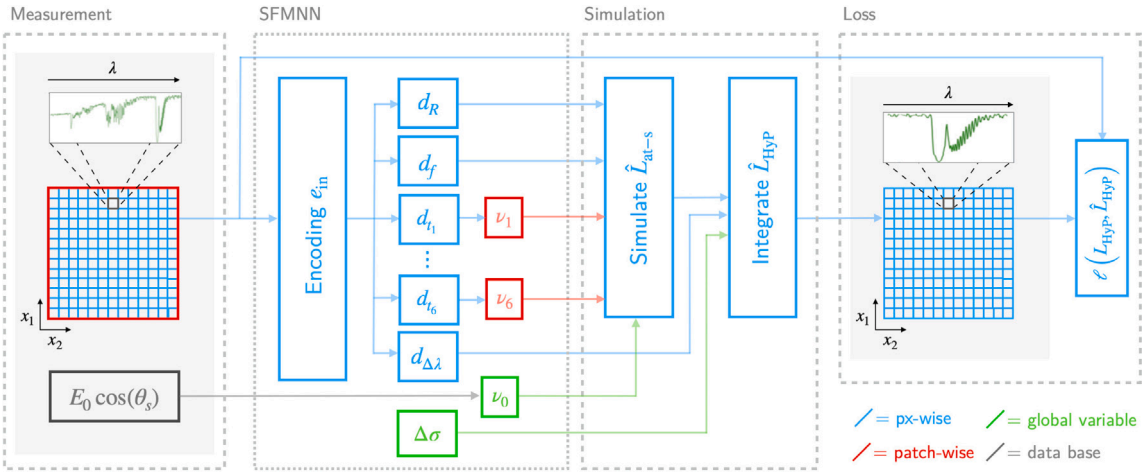


Fig. 2. Outline of SFMNN's data input and architecture. Hyperspectral image patches are encoded (e_{in}) in a pixel-wise fashion to a common latent space. The latent vectors are subsequently decoded to physical variables ($d_R, d_f, \dots$) parameterizing a four-stream model simulating the high-resolution at-sensor radiance $\hat{L}_{\text{at-s}}$. The patchwise decodings ($d_{t_1}, \dots, d_{t_6}$) are homogenized across the patch for that purpose. In a second simulation step, the estimated sensor characterization ($\Delta\lambda, \Delta\sigma$) is applied to $\hat{L}_{\text{at-s}}$ in order to yield a reconstructed observation $\hat{L}_{\text{HyP}}$. Different colors denote the range over which variables are kept constant (*px-wise* varying in every pixel, *patch-wise* over patches, *global-variable* over a single training data set). *data base* denotes the origin of the solar irradiance model (Kurucz, 2006) and solar zenith angle θ_s calculated for each acquisition. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Campaign data sets acquired in different locations in the years 2018–2023. With FLOX we denote the availability of simultaneous FLOX data, with Δh the maximum topographic variation over the campaign data set, with *HyP* the corresponding campaign report and with *DEM* corresponding the digital elevation map used for validation purposes. All campaign data sets and corresponding FLOX measurements are available (Buffat et al., 2024b) .

Campaign data set	FLOX	Δh [m]	HyP	DEM	Location
SEL-2018 (600 m)	✓	20	Rascher et al. (2021)	–	Selhausen, DE
SEL-2019 (600 m)	–	20	Rascher et al. (2022a)	–	
SEL-2022 (600 m)	–	20	–	–	
WST-2019 (1500 m)	✓	20	Rascher et al. (2022a)	–	Braccagni, IT
NRS-2019 (1500 m)	–	70	Rascher et al. (2022a)	Tarquini et al. (2023)	
TR32-2019 (1800 m)	–	150	Rascher et al. (2022a)	Nordrhein-Westfalen (2017)	Jülich, DE
INN-2022 (350 m)	–	60	–	Nordrhein-Westfalen (2017)	
SOP-2023 (600 m)	–	140	–	Nordrhein-Westfalen (2017)	
CKA-2020 (350 m)	✓	20	–	–	Klein
CKA-2020 (600 m)	✓	20	–	–	Altendorf, DE
CKA-2021 (350 m)	✓	20	–	–	
CKA-2022 (350 m)	✓	20	–	–	
GLO-2021 (1150 m)	✓	70	Rascher et al. (2022b)	NASA JPL (2013)	Mollerussa, ES

variation within single acquisitions (see *flat* in Table 2). The number of patches per finetuning data set differs depending on data availability at individual flight heights. We selected acquisitions for individual both the pretraining and finetuning data sets by hand to ensure a balanced distribution of the sun-zenith angles (SZA) at which the acquisitions were recorded.

Parallelly, we created data sets with strong topographic variability (see *topo* in Table 2). As for the *flat* denoted data sets, we grouped together campaign data sets acquired at the same flight height. Differently to D_{mix} we group the acquisitions according to the nominal flight height in the *topo* data sets. Due to data limitations, the compiled *topo* data sets are smaller and constrained to single campaign site and acquisition years in all but one case. Since the baseline iFLD and SFM implementations are not adapted to SIF retrieval under variable topography (see Section 2.3), we cannot validate the SFMNN SIF predictions in acquisitions of these datasets directly. Instead, we perform a plausibility study in Section 4.4.

An additional data set D_{day} was compiled for validation purposes. It is used both for validating SFMNN's reconstruction performance as well as the physiological plausibility of our approach in the first application study. It consists of nine repeated flights over the same location at the same nominal flight height in an irregular frequency (10:11, 10:19, 10:27, 10:34, 13:15, 13:30, 13:38, 16:11, 16:18 CEST) allowing us to validate the performance of our approach under different solar zenith angles with otherwise comparable conditions. For our analysis of the SFMNN predicted diurnal course of SIF, we leverage the universal property of SIF to be strongly correlated to PAR under non-shaded conditions. This correlation is however only expected to be fulfilled over homogeneous fields. It may be invalid only under strongly changing irradiation in pixels with complex 3D structure. Under such conditions, (i) varying sampling directions of the angular SIF emission distribution and (ii) shade impede the comparability of measurements at different times during the day. Predictions of the D_{day} data set are well suited for this analysis as the acquisitions have a relatively large pixel footprint (2.3 m) due to the large nominal flight height (1500 m). At this pixel resolution small scale inhomogeneities in crop fields are not preserved.

2.2. In-situ fluorescence measurements

A subset of the HyPlant measurement campaigns presented in Table 1 were accompanied by in-situ radiance and irradiance measurements recorded by the FLOX system (Fluorescence Box, JB-Hyperspectral Devices GmbH, Düsseldorf, Germany) (Buffat et al., 2024b). The FLOX system provides precise radiance top-of-canopy measurements within

Table 2
Compiled data sets of HyPlant acquisitions from flight campaigns in the years 2018–2023. Flight heights are not mentioned if indicated in the data set name. In the first column, we indicate the purpose of the data set: *flat* or *topo* denote the training of predictors for flat or topographically variable terrain, *val.* denotes specific validation purposes outlined in the Results section, *pretrain* denotes the pretraining of a common predictor. $|D|$ denotes the number of used 17 × 17 patches [$\times 10^3$] and in brackets the number of used acquisitions.

Data set		Included campaign data sets	$\lvert D\rvert$
pretrain	D_{mix}	SEL-2018 (600 m), NRS-2019 (1500 m), WST-2019 (1500 m), CKA-2021 (350 m), GLO-2021 (1150 m)	260 (44)
flat	D_{350}	CKA-2020, CKA-2021, CKA-2022	183 (44)
	D_{600}	SEL-2018, SEL-2019, SEL-2020, SEL-2022, CKA-2019, CKA-2020	283 (32)
	D_{1500}	WST-2019	54 (16)
topo	D'_{600}	INN-2022, SOP-2023	107 (12)
	D'_{1150}	GLO-2021	100 (8)
	D'_{1500}	NRS-2019	27 (12)
	D'_{1800}	TR32-2019	66 (8)
val.	D_{day}	WST-2019	30 (9)

small footprints of a few square meters depending on the experimental set-up. Fluorescence estimates derived from FLOX data sets have been used in the past for independent validation of HyPlant SIF estimates (Tagliabue et al., 2019; Krämer et al., 2021).

The present study provides a systematic comparison of SFM, iFLD and SFMNN predictions from HyPlant imagery with FLOX derived iFLD SIF estimates as provided by the manufacturer's processing software. FLOX data sets from 2018–2022 were taken into account. We gained iFLD based FLOX SIF estimates with the processing software provided by the FLOX manufacturer. In order to validate with a consistent data set we disregarded FLOX measurements where the irradiance in the O_2-A band recorded at the beginning and at the end of a measurement cycle varied more than 1% because this is an indicator for changing illumination conditions. The majority of measurements were, however, taken under optimal measurement conditions so that only a negligible number of data points had to be disregarded.

Apart from the SEL-2018 FLOX data set, which was acquired with a mobile FLOX system, the FLOX geolocation was acquired by a separate RTK-GPS measurement. The geolocation for the SEL-2018 FLOX data was derived from UAV data. Due to the high temporal variability of SIF, only FLOX data within ten minutes to a corresponding HyPlant overflight was considered. When multiple FLOX measurements were available within such a ten minute time window, we used the average for comparison with the HyPlant derived predictions, and the variance to compute an uncertainty estimate of the ground SIF measurement.

To account for spatial uncertainties associated with both the geolocation of the in-situ measurements as well as the field of view of single pixels in HyPlant FLUO data, we averaged the SIF prediction in a radius of 1 m, 2 m and 3.4 m for acquisitions acquired 350, 600 and 1150 m above ground level (2 pixel radius), respectively, and used the variance within this region as an uncertainty estimate for the prediction. In addition, we used the variance of multiple FLOX derived SIF measurements collected within the chosen time window of ten minutes as an uncertainty estimate. Both the uncertainties determined for the FLOX measurements and the corresponding uncertainties calculated for the HyPlant derived SIF measurements were used to compute the uncertainty on the mean absolute error (MAE) leveraging first-order error propagation.

2.3. iFLD and SFM airborne SIF estimates

Both SFM and iFLD SIF estimates are generated operationally for all HyPlant campaigns in the O_2-B and O_2-A absorption bands separately. No dedicated rerun of the iFLD and SFM baseline methods was conducted for the comparisons presented in this work. The SFM implementation used for HyPlant FLUO data follows (Cogliati et al., 2019, 2015b) and is applied directly on radiometrically calibrated, but not atmospherically corrected FLUO at-sensor radiances. In this implementation, first a single atmospheric transfer function per acquisition is estimated from the spectral information in barren pixels. Subsequently, a spectral calibration is performed in each pixel to account for sensor miscalibrations. The actual fluorescence estimation is then implemented as a least-squares minimization of the physical radiative transfer model that we have adopted in this contribution as well. In this minimization the model is fitted to radiometrically and spectrally calibrated FLUO at-sensor radiance measurements. It is noteworthy that this implementation of SFM incorporates the atmospheric transfer functions in the at-sensor radiance simulation instead of prior atmospheric correction of the FLUO measurements. A more detailed description of the implementation can be found in Siegmann et al. (2019).

The iFLD implementation for HyPlant FLUO data follows Wieneke et al. (2016) and Damm et al. (2014). Specific implementation details are reported in Siegmann et al. (2019). As the SFM implementation detailed above, this iFLD implementation is applied to radiometrically calibrated, but not atmospherically corrected HyPlant FLUO radiance measurements. A single atmospheric characterization derived with MODTRAN5 is integrated into a set of equations that can be solved analytically. Additionally, the implementation includes an empirical correction term to account for spectral miscalibrations.

3. Development of a novel spectral fitting method

We propose a novel neural-network-based method to predict SIF in the O_2-A absorption band at 760 nm. This approach could principally be adapted to other spectral domains with sensitivity to SIF as well, e.g., at HyPlant spectral resolution the O_2-B absorption band is a-priori suitable as well. The formulation in this contribution focuses on a single contiguous spectral domain (750–770 nm), however, to evaluate the feasibility of the proposed method in a restrained and controlled problem setting where simple modeling the surface reflectance and fluorescence are possible.

The proposed methodology can be regarded as a spectral fitting method (Meroni et al., 2010; Chang et al., 2020) and we call it spectral fitting method neural network (SFMNN), accordingly. Similarly to other spectral fitting methods we first construct an explicit physical model of the measured at-sensor radiance which we then fit to HyPlant at-sensor radiance (L_{HyP}). The prediction of top-of-canopy SIF at 760 nm can be directly derived from the best estimates of the parameters modeling the top-of-canopy fluorescence emission curve.

In contrast to typical spectral fitting methods such as SFM (Cogliati et al., 2015b), where a least-squares minimization is performed, we implement a feature based optimization to fit the model to the observations. Specifically, we train a neural network (see Fig. 2) in a self-supervised fashion using the objective function that is minimized in SFM as a part of the loss formulation. Self-supervised learning means that we do not need SIF ground truth for every training sample, as would be needed for supervised learning. Instead, we use the predicted SIF (as well as predicted reflectance and atmospheric transfer functions) in a physical forward simulation model (the same as used in SFM) to reconstruct the signal. Comparing the reconstructed signal to the input signal allows to evaluate a signal reconstruction error as loss for the usual backpropagation during training of the SFMNN. This loss is complemented with regularizers to ensure a physical disentanglement of atmospheric and surface related contributions to the at-sensor radiance. Training is thus completely independent of any SIF ground

measurements; a fact allowing us to finetune the model to any new data even during prediction where needed or desired.

Differently to SFM, in SFMNN atmospheric transfer functions are estimated locally over small patches which allows to account for changing observation and atmospheric conditions in a straightforward fashion. Moreover, SFMNN relies only on the full HyPlant at-sensor radiance spectrum and a general distribution of high-resolution atmospheric transfer functions. This distribution is gathered from MODTRAN5 atmospheric transfer estimates produced as a side product of SFM applied to a number of HyPlant acquisitions.

3.1. Simulation of HyPlant at-sensor radiance

The loss formulation used for training relies on an explicit forward simulation model of the measured at-sensor signal L_{HyP}, measured over the spectral range $\mathcal{W}_{\mathrm{tot}}$ (670–780 nm), in the spectral subset $\mathcal{W} \subset \mathcal{W}_{\mathrm{tot}}$. The simulation range $\mathcal{W}$ was fixed as the spectral range 750–770 nm covering the O_2-A absorption band. In order to formulate a model for the physical at-sensor signal, we make use of the four-stream forward simulation formulation used in the SFM baseline (Cogliati et al., 2015b), which is based on the Soil Canopy Observation Photosynthesis Energy balance (SCOPE) (van der Tol et al., 2009) radiative transfer model explicitly accounting for diffuse and direct fluxes. It reads

$$\begin{aligned} L_{\mathrm{at-s}} = {} & \frac{E_s^0 \cos(\theta_s)}{\pi} \\ & \times \left(\rho_{so} + \tau_{oo} \left(\tau_{ss}\, r_{so} + \frac{\tau_{sd} + \tau_{ss}\, \overline{r_{sd}}\, \rho_{dd}}{1 - \overline{r_{dd}}\, \rho_{dd}}\, r_{do} \right) + \tau_{do}\, \frac{\tau_{sd}\, \overline{r_{dd}} + \tau_{ss}\, \overline{r_{sd}}}{1 - \overline{r_{dd}}\, \rho_{dd}} \right) \\ & + \tau_{oo}\, \frac{\left(f + \overline{f_d}\, \rho_{dd}\, r_{do} \right) + \overline{f_d}\, \tau_{do}}{1 - \overline{r_{dd}}\, \rho_{dd}}. \end{aligned} \tag{1}$$

We define all quantities involved in Eq. (1) in Table 3. After rewriting the functions related to atmospheric transfer as $t_0 \dots t_6$ and products thereof as $t_7 \dots t_{12}$ (see Table 3) and simplifying Eq. (1) by setting $r_{sd} = r_{do} = R$, $\overline{r_{sd}} = \overline{r_{dd}} = \overline{R}$, $\overline{f_d} = \overline{f}$ we get

$$\begin{aligned} L_{\mathrm{at-s}} &= L^R_{\mathrm{at-s}} + L^f_{\mathrm{at-s}} \\ &= t_0 \left(t_1 + t_7\, R + \frac{t_8\, R + \overline{R}\, R\, t_{12} + \overline{R}\, (t_9 + t_{10})}{1 - t_2\, \overline{R}} \right) \\ &\quad + \left(t_5\, f_s + \frac{\overline{f}\, (t_6 + t_{11}\, R)}{1 - t_2\, \overline{R}} \right) \end{aligned} \tag{2}$$

where the overline denotes a spatial mean. In practice, this mean is computed by taking the mean over the output window. We estimate $t_1 \dots t_6$, f and R at a high spectral resolution ($\delta\lambda = 0.0055$ nm) by modeling each of these functions in dependency of the input data. The implementation details of the modeling of each of these functions is detailed below in Section 3.2.1.

For the application of the neural network to the HyPlant radiance product L_{HyP}, the simulation model needs to take into account HyPlant's sensor characteristics and potential miscalibrations in addition to the physical process which only yields $L_{\mathrm{at-s}}$. Spectral miscalibration, resulting in smile effects, can lead to a misalignment spectral features in the modeled solar irradiance and atmospheric characterization with respect to the measurements. This can lead to large uninformative gradients in the loss computation described in Section 3.4. In order to simulate L_{HyP}, we need to take into account possible sensor miscalibrations by estimating the shift $\Delta\lambda$ with respect to the nominal wavelength λ and the shift $\Delta\sigma$ with respect to the nominal spectral resolution σ. Optimally, we would consider spectral shifts in each spectral band. Considering the computational overhead resulting from the large number of spectral HyPlant bands that would need to be fitted simultaneously (172 bands), we have restricted the shift prediction. We estimate a single scalar shift $\Delta\lambda$ per spatial pixel and a single shift $\Delta\sigma$ per training set. These shifts do not correspond directly to the physical sensor miscalibration but allow to take into account the effects of spectral miscalibration that otherwise would result in large, uninformative gradients of the loss. Given an estimated center wavelength shift per pixel $\Delta\lambda$, spectral resolution σ and estimated shift in spectral resolution $\Delta\sigma$, a simple sensor model can be created

$$\hat{L}_{\mathrm{HyP}}(\lambda, x, y) = \int_{\mathcal{W}} d\lambda\; g\left(\lambda + (\Delta\lambda)_{x,y} \mid \lambda_i,\, \sigma + \Delta\sigma\right)\; L_{\mathrm{at-s}}\left(\lambda + (\Delta\lambda)_{x,y}, x, y\right) \tag{3}$$

This model gives us the at-sensor radiance as measured by HyPlant $\hat{L}_{\mathrm{HyP}}(\lambda)$ in the spectral range $\mathcal{W}$ (750–770 nm) considered in the simulation. $\hat{L}_{\mathrm{HyP}}$ can be compared to radiometrically calibrated at-sensor radiance $L_{\mathrm{HyP}}(\lambda_i)$ in the nominal center wavelengths λ_i. In practice, we assume a Gaussian Spectral Response Function (SRF) g for all λ_i. This allows us to implement Eq. (3) approximately by Gaussian smoothing followed by a linear interpolation to sample the spectra at nominal HyPlant wavelengths.

Table 3
Definitions for all quantities involved in the simulation model as defined in Eq. (1) and (2).

Eq. (1)	Eq. (2)	Definition
$E_0 \cos(\theta_s)$	t_0	solar irradiance modulated by the local solar zenith angle
f		directional fluorescence emission
$\overline{f_d}$	$\overline{f}$	diffuse fluorescence emission
ρ_{so}	t_1	bi-directional atmospheric reflectance
ρ_{dd}	t_2	spherical albedo at the bottom of the atmosphere
τ_{oo}	t_3	directional atmospheric transmittance along observation direction
τ_{ss}	t_4	directional atmospheric transmittance in solar direction
τ_{do}	t_5	directional atmospheric transmittance in observation direction for diffuse incidence
τ_{sd}	t_6	diffuse transmittance for solar incidence
r_{so}	R	bi-directional reflectance factor (BRF)
r_{do}	R	hemispheric-directional reflectance factor (HDRF)
$\overline{r_{dd}}$	$\overline{R}$	bi-hemispherical reflectance factor (BHRF) of the surrounding
$\overline{r_{sd}}$	$\overline{R}$	directional-hemispherical reflectance factor (DHRF) of the surrounding
$\tau_{ss}\, \tau_{oo}$	t_7	multiplication definitions in Eq. (2)
$\tau_{sd}\, \tau_{oo}$	t_8	
$\tau_{ss}\, \tau_{do}$	t_9	
$\tau_{ss}\, \rho_{dd}$	t_{10}	
$\tau_{oo}\, \rho_{dd}$	t_{11}	
$\tau_{ss}\, \rho_{dd}\, \tau_{oo}$	t_{12}	

3.2. Architecture

SFMNN acts on patches of spatially contiguous spectral input $c_{xy} \in \mathbb{R}^{p \times p \times (\Lambda + N_p)}$, where p denotes the spatial patch dimension, $\Lambda = 1024$ the spectral dimensionality of the full spectral range of the FLUO sensor $\mathcal{W}_{\mathrm{tot}}$ and $N_p = 3$ the number of additional variables passed to the network (nominal flight height, solar zenith angle, across-track position). We randomly draw these patches from all the acquisitions in the data set during training. The patch's dimensions p were fixed in all experiments in this work. Preliminary experiments have shown patches of 17×17 pixels to allow training batch sizes that were sufficiently large for efficient training and to constrain the optimization. While

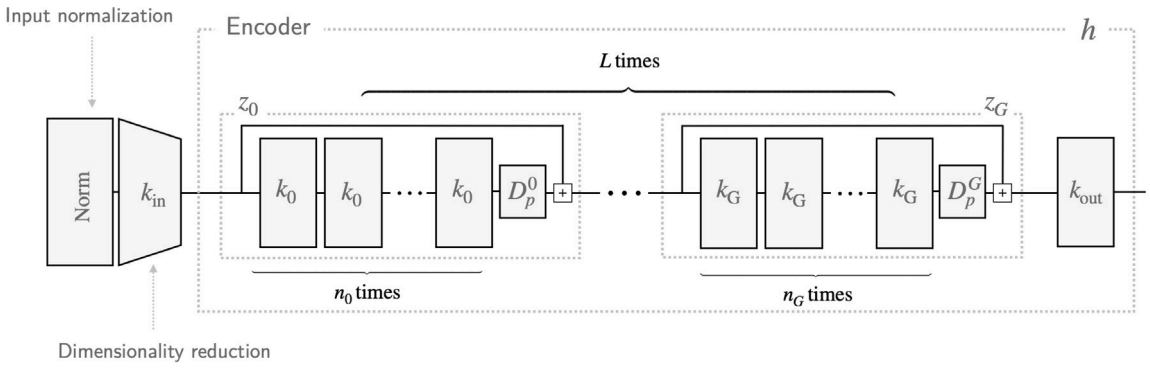


Fig. 3. Conceptual sketch of the concatenation of input normalization layer, dimensionality reduction and encoder architecture h.

we have fixed the patch size in all experiments, further investigations should consider varying this patch size as a function of the physical pixel size, i.e. the sensor flight height. This input consists of the full spectral range of the FLUO at-sensor radiance where each band is interpreted as a channel to the neural network. In addition to the spectral input we also add the nominal flight height, the solar zenith angle and the spatial across-track position as additional layers to the input patch. All parameters are predicted in a pixelwise fashion except parameters of the atmospheric functions t_i and $\Delta\sigma$. The atmospheric functions t_i are estimated on a patchwise basis to enforce the empirical expectation of large auto-correlation distances of atmospheric parameters. $\Delta\sigma$ was estimated directly as a network parameter since preliminary experiments have shown that formulating it as a function of the spectral input would affect disadvantageously the decomposition capacity of the network. SFMNN consists of a normalization layer, a dimensionality reduction layer k_{in}, an encoder e_{in} and several decoder modules for each of the variables. Both e_{in} and the decoder modules are simple multi-layer perceptrons with residual links.

The spatially dependent input c_{xy} passed to SFMNN is first normalized as presented in Fig. 3, where we show the first layers of SFMNN. The normalization layer is implemented as a batch normalization layer that is trained with the network during one epoch. In order to decrease the large input dimensionality of 1027 channels (1024 spectral channels and 3 additional variables), we introduce a dimensionality reduction to 100 channels implemented as a linear layer followed by a rectified linear unit (ReLU) activation (k_{in} in Fig. 3).

The construction of the prediction network involves transforms k_i simply defined as

$$k_{\text{i}} : \text{ReLU} \circ \text{BN} \circ \ell_{k_{i-1}\to k_i} \tag{4}$$

where $\ell_{k_{i-1}\to k_i}$ is a linear layer with input and output dimensions d_{i-1} and d_i, BN denotes a batch normalization layer (Ioffe and Szegedy, 2015; Wu and He, 2018) and ReLU a rectified linear unit. Both encoder e_{in} and decoder modules are defined as h modules with differing layer dimensions listed in Table 4. We define a module h as $k_{\text{out}} \circ z_L \circ \ldots \circ z_0$ with

$$z_i(x) = \text{D}_p^i \circ \text{BN} \circ \text{ReLU} \circ \left(\ell_{k_{i-1}\to k_i}(x) + \left(k_i \circ \ldots \circ k_i\right)(x)\right) \tag{5}$$

implementing the repeated application of layers k_i and a residual link. D_p^i denotes a dropout layer with dropout probability p. k_i are defined as in Eq. (4). The dimensions of the linear layers of all submodules of e_{in} are listed in Table 4.

3.2.1. Mapping the decoded input spectra to physical parameters

Decoder modules are constructed identically to the encoder e_{in}, i.e. they are implemented as h modules but with submodules of different dimensions as compared to e_{in}. We list all the dimensions of the decoder module in Table 4. All decoders act on the same spatially dependent encoding $p_{xy} = e_{\text{in}}(c_{xy})$ of the input c_{xy}. Each decoding $q_v = d_v(p_{xy})$ for variable v is then mapped to the physical parameters in Eq. (2). The mapping between the decoder outputs q_v and the physical values of variables that are predicted in a pixelwise fashion is implemented by a scaling layer

$$m(q \mid b_\ell, b_u) = b_\ell + \text{sigm}(q)\left(b_u - b_\ell\right) \tag{6}$$

where sigm denotes the sigmoid function. This formulation restricts the solution space to predefined parameter ranges (b_ℓ, b_u). These parameter ranges were chosen according to prior knowledge and are listed in Table 5.

We allow scalar and spatially variable $\Delta\lambda$ across acquisitions and different campaigns. We include this variation in the forward simulation by constructing a decoder $d_{\Delta\lambda}$ with a scaling layer

$$d_{\Delta\lambda}(p) = m\left(d(p) \mid \left(b_{\Delta\lambda}\right)_\ell, \left(b_{\Delta\lambda}\right)_u\right) \tag{7}$$

Similarly, we construct the decoder $d_f(p) \to (A_f, \sigma_f)$ that yields the parameterization for the top-of-canopy fluorescence emission. Since the present contribution is restricted to SIF estimation in the O_2-A absorption band, we restrict the model to the emission peak at 737 nm. We assume a normal distribution for

$$f(\lambda) = A_f\, \mathcal{N}\left(\lambda \mid \mu_\lambda, \sigma_f\right),\ \lambda \in \mathcal{W} \tag{8}$$

around the emission peak $\mu_f = 737$ nm similarly to Pato et al. (2023), Subhash and Mohanan (1997). While more precise functional expressions for top-of-canopy fluorescence have been proposed (Cogliati et al.,

Table 4
Architecture dimensionalities according to Fig. 3. Each element in a tuple denotes an architecture parameter for the ith encoder or decoder layer ($0 \le i \le G$). k_i denotes the dimensionality of the ith encoder (decoder) sublayers, n_i denotes the number of repetitions of sublayers, D_p^i denotes the dropout rate of the output of the ith encoder (decoder) layer.

	k_i	n_i	D_p^i
Encoder e_{in}	(100, 50, 50)	(3, 3, 1)	(0.1, 0)
Decoder d_v	(100, 50, 50, 50)	(3, 1, 1, 1)	(0, 0, 0, 0)

Table 5
Upper and lower bounds parameterizing the scaling layers for each variable needed in the signal reconstruction. ρ, s_ρ and e parameterize the reflectance (see Eq. (9)), A_f and σ_f parameterize the fluorescence signal model (see Eq. (8)), $\Delta\lambda$ denotes the maximum wavelength shift per band (see. Eq. (7)), E and t denote the maximum range of the estimated atmospheric transfer PCA weights (see Eq. (10)) and the solar irradiance variation (see Eq. (11)).

	ρ	s_ρ	e	A_f	σ_f	$\Delta\lambda$	E	t
Lower bound b_ℓ	0	0	0	0	20	−0.09	0.98	−300
Upper bound b_u	0.6	12×10^{-3}	1	8	21	0.09	1.02	300

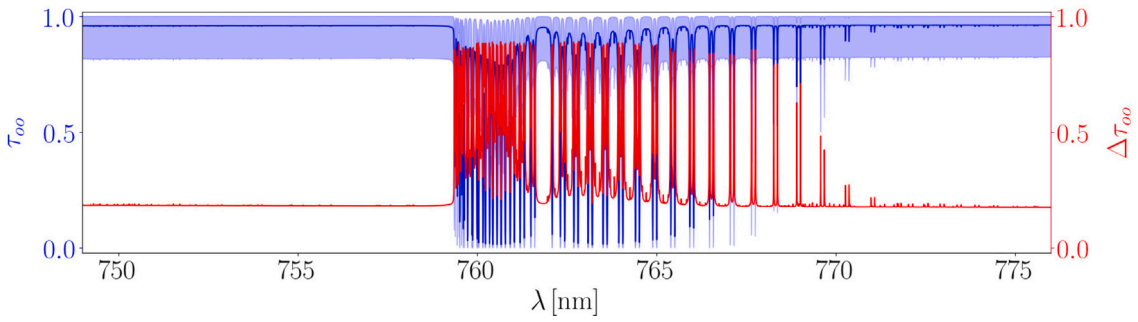


Fig. 4. Distribution of the SFM estimated transfer function τ_{oo} used for the PCA decomposition in shaded blue. We show the lower and upper bound as well as the mean τ_{oo} . The total extent of $\Delta\tau_{oo} = \max(\tau_{oo}) - \min(\tau_{oo})$ in the selected data set is shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2015b; Magney et al., 2019) we adopt here the normal distribution for simplicity. We expect the errors incurred from the model mismatch to be marginal as the spectral reconstruction is restricted to a narrow spectral range $\mathcal{W}$. The prediction bounds parameterizing the scaling layer are given in Table 5.

Equivalently, we restrict the definition of the reflectance R to a region covering tightly the simulated spectral range around the O_2-A absorption band $\mathcal{W}$. We assume R can be approximately expressed by a square polynomial in this spectral range. Following (Pato et al., 2023, 2024) a network $d_R(p) \rightarrow (\rho, s_\rho, e)$ is constructed in a way that reflectance estimates can be written

$$R(\lambda \mid p) = \rho + s_\rho(\lambda - \lambda_0) + \frac{s_\rho(e-1)(\lambda-\lambda_0)^2}{2(\lambda_1 - \lambda_0)}, \ \lambda \in \mathcal{W} \tag{9}$$

where we define $\lambda_0 = 740$ nm and $\lambda_1 = 780$ nm. The bounds of the scaling layer for ρ, s_ρ and e are given in Table 5. It has proven to allow a physiologically consistent parameterization of observed reflectance spectra.

Lastly, we highlight that we estimate a single FWHM shift $\Delta\sigma$ for the whole training data set, contrarily to the CW shifts $\Delta\lambda$ and fluorescence and reflectance parameters which are estimated for each pixel. $\Delta\sigma$ is estimated as a single free parameter in the network. We chose this simplification of the retrieval problem as preliminary analysis has shown that any larger degree of freedom in the prediction of $\Delta\sigma$ impacted the fluorescence retrieval negatively.

3.2.2. Atmosphere estimation

We predict the atmospheric functions $t_1, \ldots, t_6$ and calculate the products $t_7, \ldots, t_{12}$ defined in Eq. (2). Differently to SFM, we do not rely directly on MODTRAN simulations to estimate the atmospheric transfer. Instead, we perform a Principal Component Analysis (PCA) decomposition on a chosen set of MODTRAN simulated atmospheric functions that were derived by SFM for a range of different HyPlant acquisitions (see Fig. 4). Since HyPlant flights are operated under comparable weather conditions, we assume that we can find fitting atmospheric functions for arbitrary HyPlant acquisitions in the span of the first five PCA components. This procedure is similar to Joiner et al. (2013, 2016) where atmospheric transfer estimates are derived from linear interpolations over a decomposition of the expected space of possible atmospheric transfer functions.

In order to derive a suitable PCA decomposition we used high resolution (0.0055 nm) atmospheric transfer functions calculated during SFM optimization for 38 acquisitions from a compilation of acquisitions from 2018 and 2019 (Rascher et al., 2021, 2022a). Fig. 4 shows as an example the distribution of the SFM estimated transfer function τ_{oo} in these selected acquisitions. These acquisitions were chosen to cover a multitude of flight heights, flight directions, and day times to ensure that the span of the PCA components covers all acquisitions considered in this work. We show the distribution of atmospheric conditions for these transfer functions in Fig. 5(a) and (b).

We highlight that we included atmospheric transfer functions of SEL-2018 and WST-2019 acquisitions, which are also used for training and validation (see Table 1) in this data set. In the case of acquisitions from these two campaigns we thus guarantee that the SFM solution for the atmospheric transfer functions is in the solution space of SFMNN's atmospheric predictor.

We derived the first five principal components g_{ij} from all distributions of $t_1, \ldots, t_6$ before training the network. These components g_{ij} (with $1 \leq i \leq 6$ and $0 \leq j \leq 5$) and according means μ_i and standard deviations σ_i were then used for all experiments. In order to estimate PCA weights from the encoded HyPlant radiance input p, we construct networks $d_{t_i}(p) \rightarrow \{q_{ij}\}_{j \leq N_c}$ such that

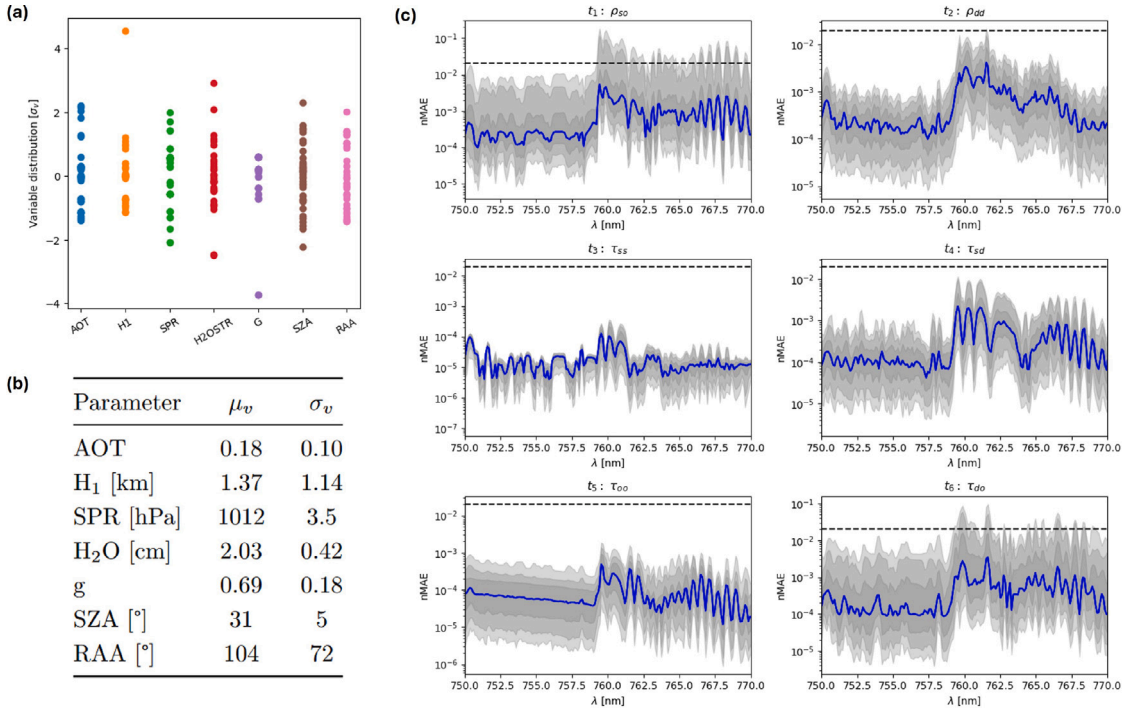
$$t_i(p) = \min\left(1, \mu_i + \sigma_i \sum_{j \leq N_c} g_{ij} w_{ij}\right), \text{ with } w_{ij} = m\left(q_{ij} \mid (b_t)_{\ell}, (b_t)_u\right) \tag{10}$$

where p is the single pixel encoding as defined above and $N_c = 5$ is the number of components used for the PCA reconstruction.

To test our assumption that atmospheric functions under typical HyPlant operation conditions may indeed be approximated by a PCA reconstruction, we show in Fig. 5 the results of a k-fold cross-validation where we evaluate the reconstruction error of each atmospheric function $t_1, \ldots, t_6$ in HyPlant spectral resolution. The reconstruction is performed using the first five components derived from 50% subsets. We draw these subsets randomly 50 times and establish optimal reconstruction weights using a least-squares minimization of the absolute error on the remaining 30% of the atmospheric functions at our disposal. From this analysis it becomes clear that almost all transfer function have a weight representation with approximation errors of less than 2% in terms of normalized absolute errors (nMAE). The reconstructions of ρ_{so} and, to a lesser degree, of τ_{do} yield nMAE 2%–10%. These approximation errors make our method less precise. Particularly, diffuse conditions may not be well represented by our atmospheric model as a consequence.

Optimally a larger data base of atmospheric transfer functions would be chosen to derive suitable PCA components to the atmospheric part of our model. We show in the following results based on a PCA decomposition derived from this small data set of 38 sets of atmospheric transfer functions and can show that it is sufficient to yield comparable SIF retrieval performance to SFM and iFLD on our in-situ validation data set.

Finally, for the solar top-of-atmosphere (TOA) irradiance t_0 we predict a single weight q_0 for the whole training dataset and use the solar irradiance spectrum E_s^0 (Kurucz, 2006) modulated by the true



(b)

Parameter	μ_v	σ_v
AOT	0.18	0.10
H_1 [km]	1.37	1.14
SPR [hPa]	1012	3.5
H_2O [cm]	2.03	0.42
g	0.69	0.18
SZA [°]	31	5
RAA [°]	104	72

Fig. 5. (a) and (b): Distribution of atmospheric conditions of atmospheric transfer functions used to derive the PCA-based atmospheric model. *AOT* denotes the aerosol optical thickness at 550 nm, H_1 the flight height above sea level, *SPR* the atmospheric surface pressure, H_2O water vapor content, g the scattering anisotropy factor, *SZA* the solar zenith angle and *RAA* the relative azimuth angle between sun and sensor. **(c):** K-fold cross-validation results of the PCA-reconstruction. We plot the normalized mean absolute error (nMAE) between target and reconstructed transfer function. Gray scales denote the 5%–95%, 10%–90% and 25%–75% percentiles and blue the median. Dashed lines denote 2% as a reference. For variable definitions see Table 3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

solar zenith angle θ_s corresponding to the acquisition time and location as initial guess for the solar irradiance at acquisition time from which the network is allowed to deviate slightly, i.e.

$$t_0 = E_s^0 \cos(\theta_s)\left(\left(b_E\right)_\ell + \mathrm{sigm}(q_0)\left(\left(b_E\right)_u - \left(b_E\right)_\ell\right)\right). \tag{11}$$

Here, q_0 is a free parameter and does not depend on the spectral input p since we introduce possible deviation from the solar irradiance E_s^0 (Kurucz, 2006) only to account for (i) a possible model mismatch and (ii) miscalibration of radiometrically calibrated HyPlant FLUO radiance. Since t_0 is supposed to be constant, the bounds $\left(b_E\right)_\ell$ and $\left(b_E\right)_u$ are chosen tightly. Simultaneous training over acquisitions recorded over time frames that do not allow the assumption of a constant solar TOA irradiance would have to account for the variability of E_s^0 by using precise daily E_s^0 estimates.

3.3. Architectural constraint

The architecture of $d_{t_i}(p)$ differs from the decoder structure introduced before, because we require a single prediction t_i per patch and not per pixel. This is necessary since we architecturally constrain the ill-posed retrieval by requiring the atmospheric prediction to hold over an extended spatial domain. We implement this by first averaging a per-pixel atmosphere encoding $d_{t_i}(p)$ over the patch. $d_{t_i}(p)$ is architecturally identical to d_f and d_R. Thus, we estimate q_i as

$$q_i = \mathrm{BN} \circ \ell_{d_a \to N_c} \circ \sum_{x,y \le s} s^{-2}\, d_{t_i}(p_{xy}) \tag{12}$$

where $\ell_{d_a \to K}$ is a linear layer with input and output dimensions d_a and N_c, where s is the number of pixels per side of a square patch in across-track dimension x and along-track dimension y and where BN denotes a batch normalization layer.

3.4. Loss & regularization

We formulate the following unsupervised loss per patch for the SIF regression problem for the spectral domain $\mathcal{W}$

$$\ell(L_{\mathrm{HyP}}, \hat{L}_{\mathrm{HyP}}) = (\ell_{\mathcal{W}} + \gamma_f \ell_f + \gamma_N \ell_{\mathrm{NDVI}} + \gamma_a \ell_{\mathrm{atm}})(L_{\mathrm{HyP}}, \hat{L}_{\mathrm{HyP}}), \tag{13}$$

where γ_f, γ_N and γ_a are weighting factors that fix the relative importance of individual loss terms. These values were fixed in preliminary experiments as $\gamma_f = 5$, $\gamma_N = 10$ and $\gamma_a = 1$ as they yielded a small difference between SFMNN and SFM SIF predictions in preliminary tests with an individual acquisition (WST 16/06/2019 11:28). Varying these weights adjusts the importance of individual loss components. No thorough validation over the sensitivity of the proposed method with respect to these values has been conducted, however.

This loss implements a least-squares minimization on the measured at-sensor radiance. It is composed of the main loss contribution over the spectral range $\mathcal{W}$ (750–770 nm)

$$\ell_{\mathcal{W}} = |\mathcal{W}|^{-1}\, \mu\left(\sum_{\lambda \in \mathcal{W}} \left(L_{\mathrm{HyP}}(\lambda) - \hat{L}_{\mathrm{HyP}}(\lambda)\right)^2\right), \tag{14}$$

where μ denotes the mean over all pixels. The three additional signal ℓ_f, physiological ℓ_{NDVI} and physical ℓ_{atm} regularizers account for data-specific constraints.

3.4.1. Signal regularization

The regularizer

$$\ell_f = |\mathcal{W}|^{-1}\, \mu\left(\sum_{\lambda \in \mathcal{W}} w_f(\lambda)\left(L_{\mathrm{HyP}}(\lambda) - \hat{L}_{\mathrm{HyP}}(\lambda)\right)^2\right)_{\delta \mathrm{R} = \delta \mathrm{t_{atm}} = 0} \tag{15}$$

weights the reconstruction with a sensor and signal specific weight function w_f. w_f boosts the loss in spectral regions with high fluorescence SNR. The gradient computation of ℓ_f is set to only influence the fluorescence module d_f as the definition of w_λ assumes known reflectance and atmospheric transfer functions.

We define w_f as the Moore–Penrose solution to a simplified and linearized retrieval problem. The derivation of the linearized problem and these weights (see Appendix) result in the definition

$$w_f(\lambda) = \left\langle \frac{1}{u_\lambda^2} \frac{\sum_{\lambda'} f_{\lambda'}^2}{\sum_{\lambda'} f_{\lambda'}^2 / u_{\lambda'}^2} \right\rangle_{p(L_{\text{at-s}}, f)} \tag{16}$$

where f denotes the predicted at-sensor fluorescence, u_λ the estimated sensor variance of $L_{\text{at-s}}$ and $\langle \dots \rangle_{p(L_{\text{at-s}}, f)}$ the expectation over a representative distribution of at-sensor radiance and fluorescence emission. We show the mean weights w_f used in all experiments that we estimate using a fixed distribution $p(L_{\text{at-s}}, f)$ in Fig. 1. Both the variability of w_f as well as the influence of w_f on the learned retrieval have not been investigated in this work.

3.4.2. Physiological constraint

The regularizer

$$\ell_{\text{NDVI}} = \mu \left(\hat{f} \, \delta \left(\text{NDVI} \leq \tau_{\text{NDVI}} \right) \right) \tag{17}$$

was introduced to reinforce the physiological expectation of vanishing fluorescence in sparsely vegetated areas. In Eq. (17) the prediction network's fluorescence prediction at 760 nm, $\hat{f}$, is evaluated in pixels with low vegetation cover as expressed by a threshold on NDVI, τ_{NDVI}. We fixed $\tau_{\text{NDVI}} = 0.15$ in all experiments in accordance with a similar thresholding principle in SFM (Cogliati et al., 2019). The NDVI is derived online during training and inference from the FLUO input spectra.

3.4.3. Physical regularization

For practical purposes, we define the total effective predicted transfer function with respect to the direct solar irradiance

$$t_{\text{tot}} = \frac{L_{\text{at-s}}^R \, \overline{R}^{-1}}{E_s^0 \cos(\theta_s)} \tag{18}$$

describing both downwelling, upwelling and diffuse contributions to the at-sensor signal. In this computation, we only use the reflectance signal contribution $L_{\text{at-s}}^R$ from Eq. (2). We point out that the resulting effective transfer function t_{tot} may not respect $\max(t_{\text{tot}}) < 1$ because of diffuse contributions to the at-sensor signal even though individual components $t_1 \dots t_6$ are constrained to < 1. This is caused by the definition of t_{tot} which relates only to the direct solar irradiance. The additional diffuse downwelling and upwelling contributions in the simulation model used in this work are not part of the normalizing denominator and can, thus, lead to $t_{\text{tot}} > 1$. We can assume the diffuse components to the measured at-sensor radiance to be small under typical HyPlant observation conditions. We make sure that the atmospheric transfer functions result in a physically plausible total transfer function t_{tot} that is not much larger than 1. by regularizing the loss with a soft constraint to enforce $\text{ReLU}(t_{\text{tot}} - 1)$ to be small.. Thus, we add

$$\ell_{\text{atm}} = \text{ReLU}(t_{\text{tot}} - 1) \tag{19}$$

as an additional regularizer to the loss. It constrains the network to decrease diffuse parts of the predicted atmospheric transfer. While this regularizer achieves the soft constraint with a suitable weight γ_a in the loss (see Eq. (13)), it has the downside that the fixed hyperparameter γ_a effectively controls the contribution of simulated diffuse at-sensor radiance.

3.5. Training set up

The training of the SFMNN network was conducted in two phases. First we performed a pretraining on a data compilation of acquisitions from different nominal heights $\mathcal{D}_{\text{mix}}$ (cf. Table 1). This pretraining provided us with a coarsely trained network p_{mix}, which we then fine-tuned to gain predictors for specific observation and terrain conditions. We trained networks p_{350}, p_{600} and p_{1500} for individual nominal flight heights (350, 600 and 1500 m) in flat terrain. In Table 1 we detail the composition of the corresponding compilations $\mathcal{D}_{350} \dots \mathcal{D}_{1500}$ that were used for fine tuning. In addition to the flat terrain predictors, we also compiled data sets with variable topography ($\mathcal{D}_{600}^t$, $\mathcal{D}_{1150}^t$, $\mathcal{D}_{1500}^t$ and $\mathcal{D}_{1800}^t$). Due to a significantly reduced availability of HyPlant data in topographically complex terrain these compilations mainly consist of acquisitions from a single campaign.

The HyPlant acquisitions in a training data set were cut spatially in partially overlapping patches. A fixed patch size of 17 × 17 HyPlant pixels was chosen. While we used the whole spectral range covered by the FLUO module as spectral input window $\mathcal{W}_{\text{tot}}$, we restricted the prediction range to the more constrained simulation range $\mathcal{W} = [750, 770]$ nm. A larger simulation range would have required modeling of the at-sensor radiance in a larger spectral domain. Since this work focuses on the O_2-A absorption band we instead restricted the fluorescence estimate to 750–770 nm while still allowing the network to leverage information on the fluorescence emission outside the spectral window which we simulate. In all training runs, the input provided to the network consisted of concatenated layers of the spectral information, the across-track pixel location as well as the mean solar incidence angle per acquisition. With respect to the analysis of predicted transfer functions in topographically variable terrain, we point out that no topographic elevation such as a digital elevation model (DEM) was provided to SFMNN during training or testing time.

The training was conducted with an Adam optimizer guided gradient descent Kingma and Ba (2017). The nominal learning rate was set initially to $\ell_r = 10^{-3}$ and subsequently reduced to a minimum of $\ell_r = 10^{-4}$ with a learning rate scheduler. All networks (p_{mix}, $p_{350} \dots$) were trained until the validation loss on a random subset of the training data stopped decreasing.

4. Results

In order to validate the SFMNN reconstruction and retrieval performance, we present a qualitative comparison of the reconstructed at-sensor radiance of SFM and SFMNN as well as a quantitative comparison of SFMNN top-of-canopy fluorescence estimates against top-of-canopy FLOX measurements. Additionally, we include two application studies to assess the physiological and physical plausibility of SFMNN's estimated parameters.

4.1. At-sensor radiance reconstruction performance

Good reconstruction performance of the at-sensor radiance is a critical prerequisite for successfully disentangling R, f and the atmospheric transfer functions t_i. In order for SFMNN to yield consistent SIF estimates, comparable reconstruction performance across a wide range of observational conditions is needed. While, the reconstruction is an important evaluation metric, it does, however, not guarantee the correctness of the estimated at-sensor signal decomposition. This is due to the ill-posed nature of this decomposition problem that may allow ambiguous solutions resulting in similarly small residuals.

We show that SFMNN reaches a reconstruction performance similar to SFM by analyzing error statistics of the predicted at-sensor radiance both in the full simulation range $\mathcal{W}$ (750–770 nm) as well as in the spectral range $\mathcal{W}_{O_{2-A}} = [759.55, 761.60]$ nm in the O_2-A absorption band. In Fig. 6 we plot the 5%–95% percentiles of the absolute reconstruction residuals along with the mean absolute

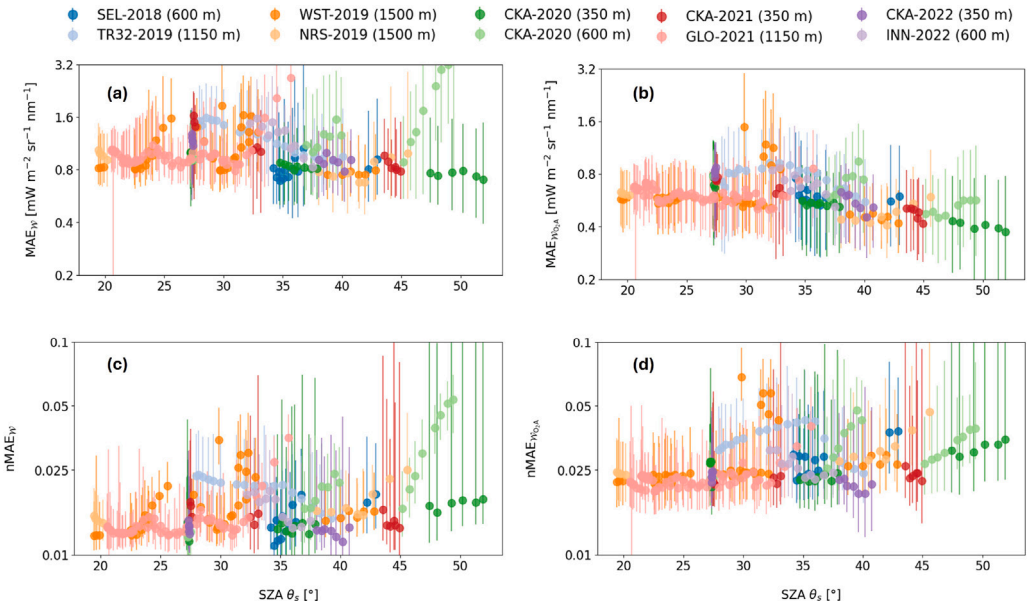


Fig. 6. Statistics of absolute errors across single acquisitions. We show the mean 5, 50 and 95% quantiles of absolute error (MAE, upper row) and of normalized absolute error (nMAE, lower row) between observed radiance and radiance predicted by SFMNN in the spectral ranges $\mathcal{W}$ = [750, 770] nm (*(a)* and *(c)*) and $\mathcal{W}_{O_2\text{-}A}$ = [759.55, 761.60] nm (*(b)* and *(d)*) as a function of the sun zenith angle (SZA). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

error (MAE). The SFMNN simulations exhibit overall mean absolute error ($MAE_{\mathcal{W}}$) distributions constrained consistently in the range 0.6–2 mW nm^{-1} sr^{-1} m^{-2} (Fig. 6(a)) and $MAE_{\mathcal{W}_{O_2A}}$ in the O_2-A band in the range 0.4–1 mW nm^{-1} sr^{-1} m^{-2} (Fig. 6(b)). We equally show the normalized mean absolute error distributions

$$nMAE_{\mathcal{W}} = \frac{\left|\hat{L}_{HyP} - L_{HyP}\right|}{L_{HyP}} \tag{20}$$

for the spectral range $\mathcal{W}$ in Fig. 6(c) and equally for $\mathcal{W}_{O_2A}$ in (d). Contrarily to the MAE, we find increasing nMAE outliers with increasing SZA. The MAE decreases more slowly than the mean at-sensor radiance with increasingly diffuse conditions for image data acquired under high SZA. As a consequence, the nMAE increases under low-light conditions.

In order to assess whether SFMNN's reconstruction error is sufficient for a spectral fitting based SIF retrieval, we compare the reconstruction residuals of SFMNN and the SFM baseline. The present study only focuses on a comparison in the D_{day} dataset. To highlight the variability in reconstruction performance across different light regimes we differentiate between *noon* acquisitions taken around the solar noon (13:15, 13:30, 13:38 CEST) and *off-noon* acquisitions taken in the morning and afternoon (10:11, 10:19, 10:27, 10:34, 16:11, 16:18 CEST).

In Fig. 7(a) and (b) we plot the mean absolute error distributions $MAE_{\mathcal{W}}$ over simulation range $\mathcal{W}$ (750–770 nm) and $MAE_{\mathcal{W}_{O_2\text{-}A}}$ in the O_2-A band (759.55–761.60 nm). SFMNN's reconstruction performance over the full spectrum (see Fig. 7(a)) is sightly worse than SFM. Contrarily, in the O_2-A band (see Fig. 7(b)) we see an improvement of SFMNN over SFM. In both cases, we find lower reconstruction errors in *off-noon* than in *noon* acquisitions, but increased nMAE in *off-noon* consistent with the discussion of Fig. 6 above. This partially improved performance of SFMNN over SFM in the O_2-A band of in the D_{day} data set can also be observed in terms of $nMAE_{O_{2-A}}$ in 7 (c) and (d). As for the absolute MAE, we find worse overall performance $nMAE_{\mathcal{W}}$ but an improvement of SFMNN in the O_2-A band.

We conclude that SFMNN is able to learn a reconstruction of the at-sensor radiance with a reconstruction error in the O_2-A band consistently bounded in the range 0.4–1 mW nm^{-1} sr^{-1} m^{-2} over the whole range of observational conditions covered by the data sets. SFMNN's prediction accuracy in the O_2-A band is improved over the operational SFM predictions in the D_{day} dataset both in terms of MAE and nMAE, but falls short of SFM's performance when comparing residuals in the full simulation range $\mathcal{W}$ in off-noon conditions.

4.2. Validation of SFMNN predicted top-of-canopy fluorescence with FLOX measurements

In order to evaluate whether SFMNN's decomposition of the at-sensor radiance yields SIF estimates comparable to SFM and iFLD baselines, we directly validate SFMNN SIF with in-situ SIF measurements that were acquired quasi-simultaneously to the airborne HyPlant acquisitions. In particular, we compare FLOX derived iFLD SIF estimates to HyPlant derived SIF predictions for the five measurement campaigns SEL-2018 (600 m), WST-2019 (1500 m), CKA-2020 (350 m), CKA-2020 (600 m), CKA-2021 (350 m), GLO-2021 (1150 m) and CKA-2022 (600 m) (see Table 2). We show comparative scatter plots (Fig. 8) and performance metrics of HyPlant derived SFM, SFMNN and iFLD (Table 6) taking into account all in-situ FLOX measurements available after filtering as described in Section 2.2. In addition to the MAE, the results report the Pearson correlation *r*.

The validation results show that SFMNN outperforms the two baseline methods in terms of *r* for all but two datasets (see Table 6). In these two data sets providing a lower validation accuracy, CKA-2020 (350 m) and CKA-2021 (350 m), *r* is still close to the best performing baseline method. A similar conclusion can be drawn when comparing the performance in terms of mean absolute error (MAE). While the MAE of SFMNN suffers from bias leading to overestimation in all datasets it ranks at least second in all but one dataset (CKA-2021 (350 m)) The overall validation results in Table 6 and the scatter plots in Fig. 8 suggest that SFMNN has the strongest sensitivity to changes in in-situ measured SIF, but that the prediction bias leads to a underperformance with respect to iFLD and SFM in terms of MAE.

The results for the CKA-2020 (350 m) and (600 m) runs refer to data acquired by three different FLOX systems placed in two different crop types (winter wheat, oat). In order to avoid biased metrics due to sensor specific and crop type phenological differences that potentially influenced the FLOX measurements, we additionally list the prediction performance over the measurement series of the individual measurement stations at the ground (see Table 7). The results in Table 7 show that there is variation in the performance metrics across the

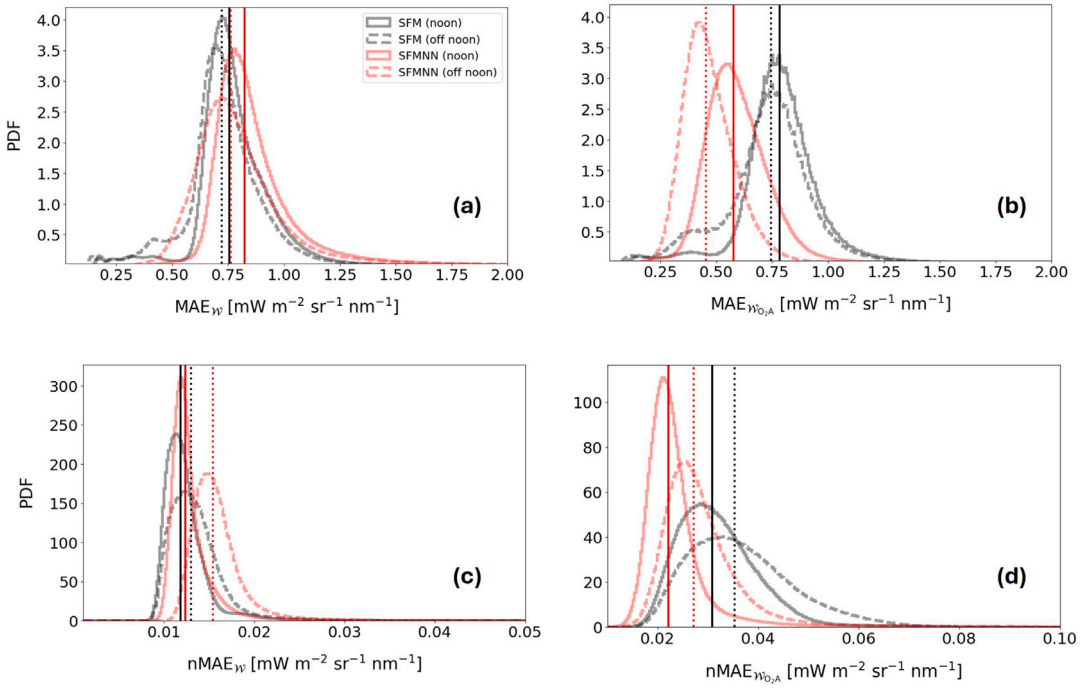


Fig. 7. Residual distribution of SFM and SFMNN reconstructions of WST-2019 (1500 m) acquisitions sampling a single day (2019/06/19). *noon* denotes data from 12:00–14:00 (3 acquisitions), *off-noon* denotes data acquired earlier than 12:00 or later than 14:00 (6 acquisitions). Shown are (a) absolute error distributions MAE$_{\mathcal{W}}$ over the complete spectral window $\mathcal{W}$, (b) MAE$_{\mathcal{W}_{O_2A}}$ in the O_2-A band [759.55, 761.60] nm and corresponding normalized absolute error distributions (c) nMAE$_{\mathcal{W}}$ and (d) nMAE$_{\mathcal{W}_{O_2A}}$ as defined in Eq. (20). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
FLOX derived SIF measurements compared to SFMNN, SFM and iFLD SIF predictions (≤10 min time difference). Correlation r marked with * have $p > 0.05$. Mean absolute errors (MAE) are given in mW nm^{-1} sr^{-1} m^{-2}. N denotes the number of validation points. We highlight the best results in bold font. Multiple MAE results are highlighted, if the uncertainty is too large to establish a single *best* case. All uncertainties are computed from the variation of HyPlant derived SIF estimates within a fixed radius (350 m: 0.5 m, 600 m: 1 m, 1150 m: 2 m) around the measurement location and the variation of FLOX SIF estimates within the predefined time window of 10 min.

Data set		r	MAE	N
SEL-2018 (600 m)	SFM	0.81	0.82 ± 0.12	9
	SFMNN	**0.96**	0.68 ± 0.08	11
	iFLD	0.78	**0.58 ± 0.09**	10
WST-2019 (1500 m)	SFM	−0.35*	0.48 ± 0.07	22
	SFMNN	**0.58**	**0.20 ± 0.08**	22
	iFLD	−0.12*	0.78 ± 0.09	18
CKA-2020 (600 m)	SFM	**0.83**	0.42 ± 0.05	23
	SFMNN	**0.83**	**0.29 ± 0.05**	23
	iFLD	0.52	0.39 ± 0.08	23
CKA-2020 (350 m)	SFM	**0.90**	0.36 ± 0.04	37
	SFMNN	0.86	0.35 ± 0.04	37
	iFLD	0.55	**0.28 ± 0.05**	36
GLO-2021 (1150 m)	SFM	0.89	**0.24 ± 0.09**	6
	SFMNN	**0.91**	0.38 ± 0.15	6
	iFLD	0.81	0.73 ± 0.14	6
CKA-2021 (350 m)	SFM	0.64*	0.44 ± 0.07	7
	SFMNN	0.82	0.65 ± 0.09	7
	iFLD	**0.87**	**0.12 ± 0.15**	7
CKA-2022 (350 m)	SFM	0.58*	**0.39 ± 0.13**	6
	SFMNN	**0.70***	**0.35 ± 0.17**	6
	iFLD	−0.87*	1.12 ± 0.21	4

Table 7
Performance metrics for different 3 different FLOX devices. We denote the different devices by the crop type the devices were pointed at. Metric definitions as in Table 6. Correlation scores r marked with * have $p > 0.05$. Mean absolute errors (MAE) are given in mW nm^{-1} sr^{-1} m^{-2}. We highlight the best results in bold font. N denotes the number of validation points.

Data		FLOX	r	MAE	N
CKA-2020 (350 m)	SFM	Wheat (1)	**0.62**	0.48 ± 0.05	11
		Wheat (2)	0.86	**0.30 ± 0.07**	14
		Oat	**0.84**	0.33 ± 0.06	12
		All	**0.90**	0.36 ± 0.04	37
	SFMNN	Wheat (1)	0.59*	**0.31 ± 0.06**	11
		Wheat (2)	**0.90**	0.49 ± 0.09	14
		Oat	0.39*	**0.23 ± 0.07**	12
		All	0.86	0.35 ± 0.04	37
	iFLD	Wheat (1)	−0.19*	**0.33 ± 0.07**	11
		Wheat (2)	0.11*	**0.34 ± 0.09**	13
		Oat	0.10*	**0.17 ± 0.09**	12
		All	0.55	**0.28 ± 0.05**	36
CKA-2020 (600 m)	SFM	Wheat (1)	**0.88***	0.85 ± 0.09	4
		Wheat (2)	0.86	**0.32 ± 0.10**	10
		Oat	0.69	0.34 ± 0.06	9
		All	**0.83**	0.42 ± 0.05	23
	SFMNN	Wheat (1)	0.81*	**0.11 ± 0.08**	4
		Wheat (2)	**0.90**	**0.31 ± 0.10**	10
		Oat	**0.73**	0.36 ± 0.06	9
		All	**0.83**	**0.29 ± 0.05**	23
	iFLD	Wheat (1)	−0.56*	0.51 ± 0.10	5
		Wheat (2)	0.24*	0.45 ± 0.16	9
		Oat	0.10*	**0.25 ± 0.10**	9
		All	0.52	0.39 ± 0.08	23

different FLOX measurement set-ups. It can be observed, however, that the best performing method in r and MAE also performs best on a majority of the individual FLOX set-ups such that the overall performance assessment in Table 6 is not driven by single set-ups in the CKA-2020 data sets. Furthermore, the variation of MAE across the different measurement set-ups is of the same order as the large uncertainties on the MAEs of individual measurement set-ups for all methods.

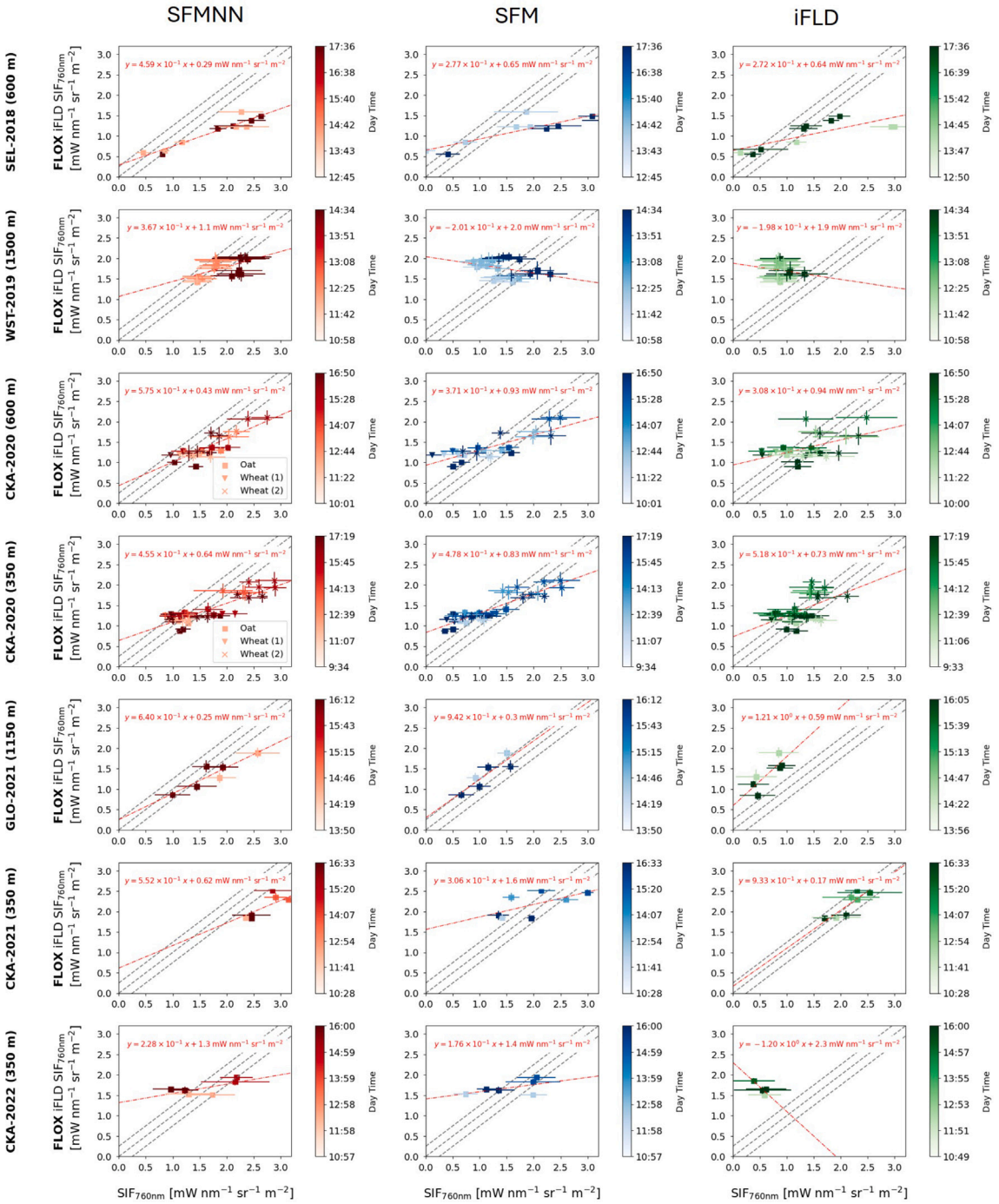


Fig. 8. FLOX derived iFLD SIF vs. HyPlant derived SFMNN, SFM and iFLD SIF in the 7 in-situ validation data sets (see Table 1). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

As a conclusion, we highlight that the SFMNN performs best in terms of r but is impacted by prediction bias. In particular, SFMNN showed sensitivity to in-situ FLOX SIF data sets where both iFLD and SFM do not (WST-2019, CKA-2022). Strong positive bias in all datasets leads to an underperformance of SFMNN with respect to iFLD and SFM in various data sets such that it only outperforms these two baseline methods in less than half of the considered validation data sets. Furthermore, a restricted comparison of the performance of individual and aggregated FLOX performance metrics in the CKA-2020 datasets indicates that iFLD, SFM and SFMNN performances can be generalized across the different measurement set-ups within single campaigns given the large uncertainties associated of both the FLOX and the HyPlant SIF estimates.

4.3. Application study 1: Retrieval of SIF in the diurnal course

In this application study we analyze the diurnal course of SFMNN predicted SIF and assess whether it peaks around the solar noon or if, contrarily, the largest fluorescence values can be found prior to or after solar noon. We formulate this plausibility test on the diurnal data set D_{day}, by investigating the overall diurnal course of predicted SIF over a large number of pixels. The SIF predictions for the nine HyPlant acquisitions in D_{day} were georegistered to form time series $f(\underline{x}, t)$ in each georegistered pixel $\underline{x}$ (see Fig. 10). We detrended each

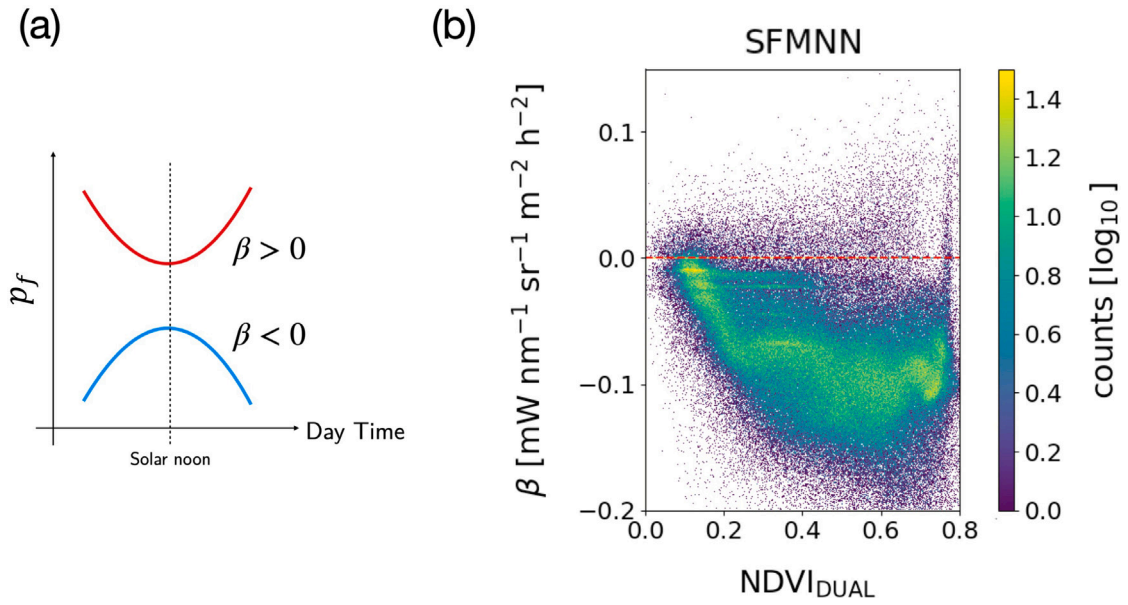


Fig. 9. (a) Sketch of two possible time series of predicted fluorescence $f(\underline{x}, t)$. Only the blue time series with $\beta < 0$ is considered plausible as it peaks during solar noon. (b) Histogram of SFMNN β-NDVI relationship shown in Fig. 10. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

time series independently and fitted a second order polynomial $p_f(\underline{x})$ to the time series in each pixel. The second order derivative of this polynomial with respect to time $\partial^2 p_f / \partial t^2 = \beta$ then was used as a proxy for the diurnal SIF dynamics (see Fig. 9(a)). In particular, we indicate whether the predicted SIF peaks around solar noon ($\beta < 0$) or in morning and evening ($\beta \geq 0$) acquisitions. In physiologically plausible SIF time series of healthy vegetation under a typical course of diurnal solar irradiance variation without shadowing, we expect $\beta < 0$ over the whole NDVI domain. The map of β for $\mathcal{D}_{\text{day}}$ in Fig. 10 highlights that pixels with $\beta > 0$ are mainly distributed parallel to tree lines where the strong shadowing in the diurnal course invalidates the model assumptions. SFMNN predictions thus yield a plausible diurnal course of SIF.

4.4. Application study 2: Retrieval of SIF in hilly terrain

In this application study we show that the constraints imposed on SFMNN's weight optimization yield topography-aware predictors. To this end, we have trained SFMNN p^t_{600}, p^t_{1150}, p^t_{1500} and p^t_{1800} on the data sets $\mathcal{D}^t_{600}$, $\mathcal{D}^t_{1150}$, $\mathcal{D}^t_{1500}$ and $\mathcal{D}^t_{1800}$ exhibiting large topographic variation. For this we compare the geometrical path length d to the fractional band depth

$$n_{O_2A} = \frac{t_{\text{tot}}(760.5\,\text{nm})}{t_{\text{tot}}(755\,\text{nm})} \tag{21}$$

predicted by SFMNN for each spatial pixel. The path length is defined as the distance computed from the surface to the sensor considering the flight height, topography and the instantaneous viewing angle at recording time.

In Fig. 11 we show an exemplary acquisition of a strong topographic gradient along a height difference of ~140 m. The SFMNN predictor adapts the predicted band depth locally to gain a similar $\text{nMAE}_{\mathcal{W}_{O2A}}$ over the full topographic change (see Figs. 11 and 12(a)). As a result we can observe a relatively constant SFMNN SIF prediction along-track for the different height ranges (see Figs. 11 and 12(b)). A small increase of maximum SIF can still be observed over the topographic change (and d decrease). It is, however, not possible to completely disassociate this SIF distribution shift from a possible change in surface conditions. This is also visible in Fig. 12(b) where we show the SIF prediction distribution as a function of the topography. Similarly, the distribution of spectral wavelength shifts $\Delta\lambda_{x,y}$ and ρ_{752} are approximately constant in the along track dimension, suggesting that the model does not compensate the atmospheric transfer over the hill slope by varying these parameters.

The predicted O_2-A band depth is not completely decoupled from surface related quantities. This can be seen in the exemplary acquisition plot in Fig. 11 in some correlated features in SIF in n_{O_2A}. We cannot determine from observational data whether this coupling is physical (i.e. conditioned by common causal factors) or conditioned by a learned empirical distribution. Furthermore, there is an unphysical artefact visible in the band depth estimation that correlates with a region of high reconstruction errors $\text{nMAE}_{\mathcal{W}_{O_2A}}$ (see Fig. 11). While we cannot pinpoint exactly, what the reason for this prediction failure is, we test for consistency of n_{O_2A} in general prediction scenarios. In order to assess whether the variation in predicted band depth n_{O_2A} is consistent across multiple acquisitions and across different predictors, in Fig. 13 we show the linear extinction factor $s_{n_{O_2A}}$ defined as the first-order derivative to the approximation

$$\mathbb{E}\left[n_{O_2A} \mid d\right] = s_{n_{O_2A}} \cdot d + \text{const.} \tag{22}$$

where d denotes the geometric path between the surface and sensor. As we are interested in the sensitivity of n_{O_2A} to the changing geometric path length, we summarize any other influencing factor under *const.* and determine $s_{n_{O_2A}}$ with a linear fit to the conditional distribution $p(n_{O_2A}|d)$ as shown in Fig. 12(a). We find that $s_{n_{O_2A}}$ stabilizes at a constant value in acquisitions with large differences in geometric path lengths $\Delta d = \max d - \min d$. This is true for acquisitions of different SFMNN models and across different data sets. As expected, the variation in the $s_{n_{O_2A}}$ estimates increases when the geometric path length differences Δd become smaller and the uncertainty on the per-acquisition expectation $\mathbb{E}\left[n_{O_2-A} \mid d\right]$ increases. As a consequence, the effect of variable d cannot be ascertained in acquisitions with small Δd. The smaller predicted band depth differences are more strongly driven by other, possibly non-physical factors. However, in acquisitions with larger Δd band depth changes are explained by a common atmospheric transfer model yielding a constant $s_{n_{O_2A}}$ (Fig. 13). This common transfer model is predicted by independent SFMNN models. We find thus a strong indication that SFMNN models converge to solutions where topographic changes are explained mainly by changing atmospheric

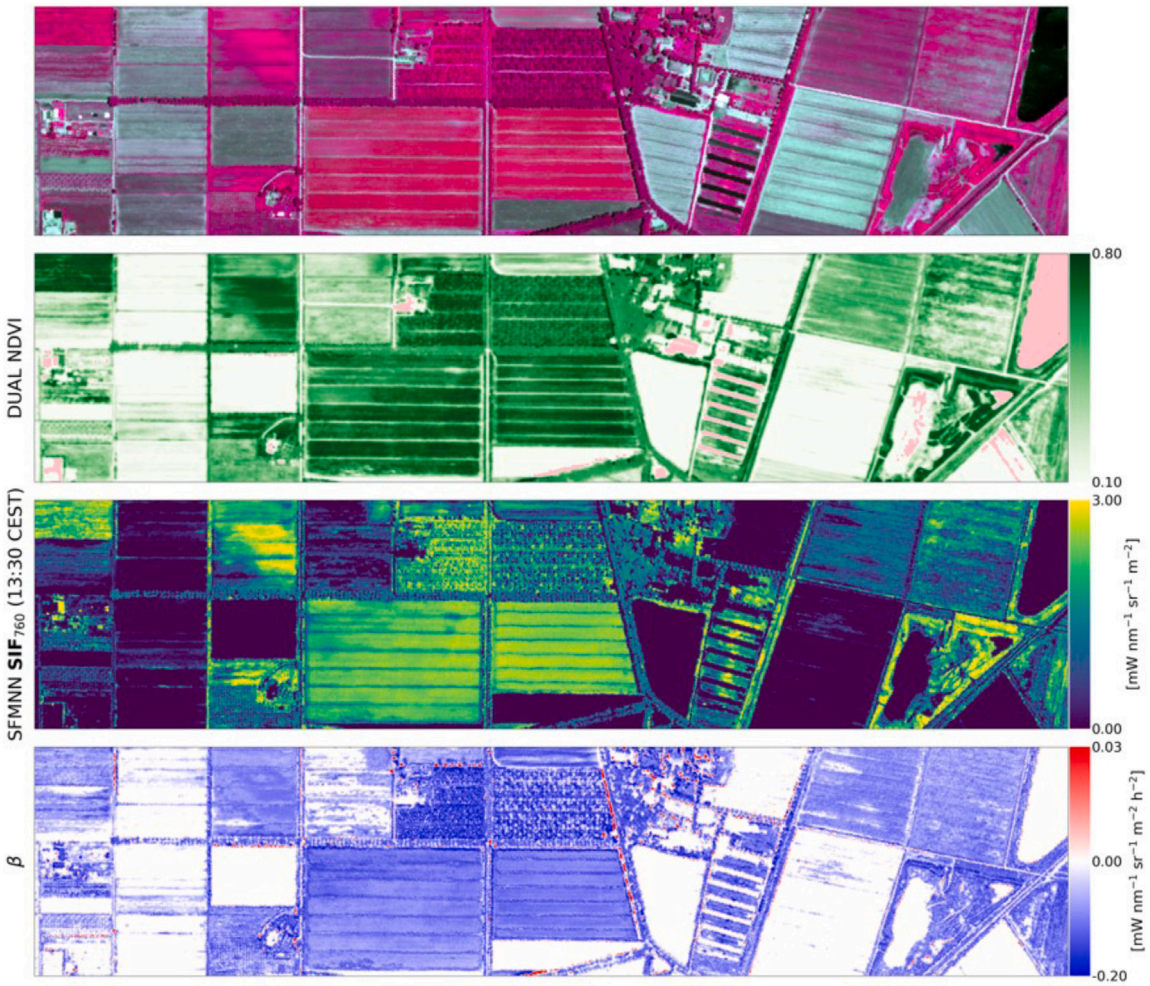


Fig. 10. We show a pseudo-color image (FLUO NIR, red and green) of the WST-2019 (1500 m) 13:30 CEST acquisition (first row), the corresponding NDVI computed from DUAL (second row), the SFMNN SIF prediction (third row) and the diurnal SFMNN SIF variation parameterized by β (fourth row). Pixels with invalid NDVI values (in shaded, barren and water pixels) are highlighted in pink. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

transfer and not, for example, by increasing fluorescence.

5. Discussion

5.1. Reconstruction performance

Since the SFM is the operational spectral fitting SIF retrieval algorithm for HyPlant data, its prediction as well as reconstruction performance has been repeatedly validated using top-of-canopy spectral measurements (Rascher et al., 2021; Cogliati et al., 2019; Siegmann et al., 2021) and cross-validated with SIF products from different platforms (Wang et al., 2022). Matching SFM's reconstruction performance is thus an important feature of novel spectral fitting algorithms aiming to retrieve SIF in addition to a good performance against other baseline methods applied to HyPlant data such as iFLD. While a good reconstruction performance does not validate the physicality of our approach, it constitutes a necessary precondition for any spectral fitting method. Failure to satisfactorily approximate the model to observations could indicate either an incomplete model formulation or too stringent constraints imposed as regularization on the fitting process.

We could show that the reconstruction performance of SFMNN is comparable to SFM in a data set of nine HyPlant acquisitions recorded over the course of a single day. SFMNN's reconstruction led to smaller MAE than SFM inside and to only slightly worse performance outside of the O_2-A absorption band. We hypothesize that the SNR-based loss weighting, which increases the importance of residuals within the absorption band, led to this preferential improvement of absolute residuals as no such prior assumption is considered in the spectral fitting process of SFM. In parallel, the improved reconstruction performance of SFMNN could be observed when analyzing its radiance normalized residuals nMAE inside the O_2-A band. We found the same differentiation in nMAE between *noon* and *off-noon* conditions for SFMNN as for SFM. Worse nMAE are found in *off-noon* conditions. This performance deterioration could also be established in SFMNN residual statistics of multiple acquisitions (cf. Fig. 6).

We have investigated the cause for this decreased reconstruction performance and find that it is likely due to lacking disentanglement of reflectance and atmospheric transfer performance in *off-noon* conditions. In Fig. 14 we show that SFMNN's reflectance prediction is not accurate in *off-noon* conditions. A significant underestimation of SFMNN as compared to georegistered DUAL reflectances can be observed in the morning and afternoon while the *noon* acquisitions show comparable reflectance distributions. In order to reduce the reconstruction error the network predictor instead increases the diffuse contributions to the at-sensor radiance in *off-noon* predictions as can be seen in Fig. 15, where we show the mean nadir t_{tot} in all $\mathcal{D}_{day}$ (a) and the simulated t_{tot} in black (b). The simulation is performed by (i) simulating L^{at-s} using a MODTRAN derived radiative transfer emulator Pato et al., 2024) and (ii) normalizing with the solar irradiance and reflectance model used in this work. We parameterize the simulation of

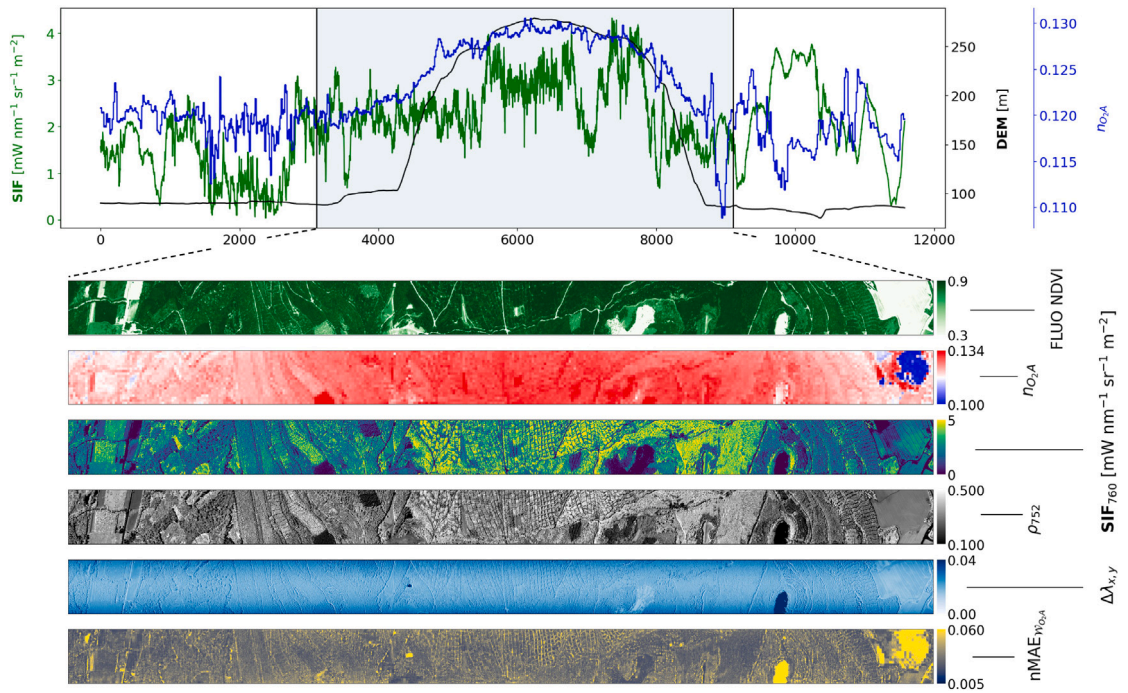


Fig. 11. An exemplary airborne data acquisition with pronounced topographic variation (SOP-2023, 2023/06/23 14:44 CEST). In the first row we show SFMNN SIF prediction (green), a digital elevation map (DEM) of the scene (black) and the predicted n_{O_2A} (blue) averaged over the across-track dimension. In the second to fourth row we show maps of FLUO derived NDVI, n_{O_2A}, SFMNN SIF SFMNN predicted reflectance at 740 nm, SFMNN predicted wavelength shifts $\Delta\lambda_{x,y}$ and reconstruction errors nMAE$_{\mathcal{W}_{O_2A}}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

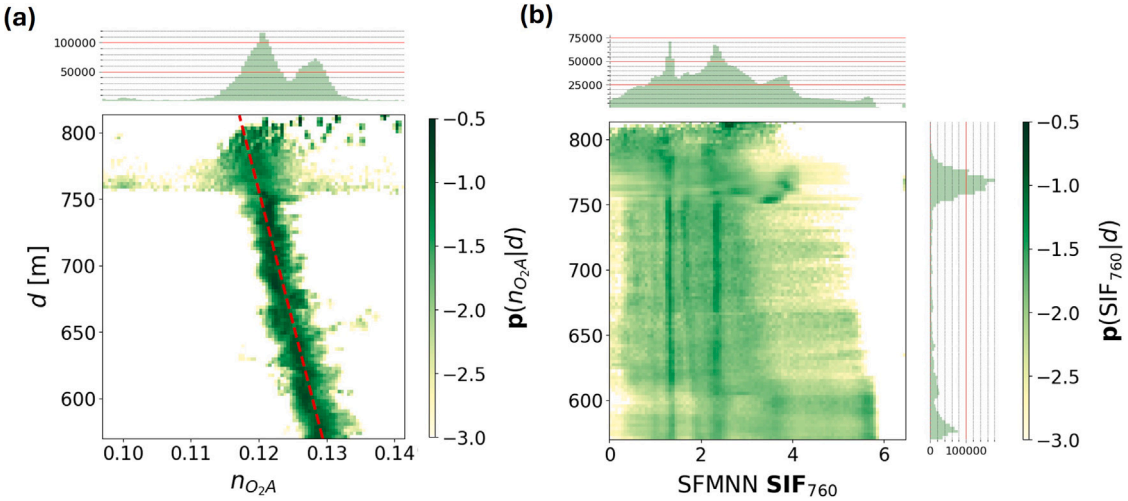


Fig. 12. **(a)**: Distribution $p(n_{O_2A}|d)$ (O_2-A band depth as a function of geometric path length d) (green) and linear fit (red) of the acquisition shown in Fig. 11 (SOP-2023, 2023/06/23 14:44 CEST). **(b)**: Distribution $p(\mathrm{SIF}_{760}|d)$ (SFMNN SIF as a function of geometric path length d). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

L_{HyP} with the mean values assumed for $\mathcal{D}_{\mathrm{day}}$ *noon* acquisitions. We also show simulations under shifts $\Delta\rho$ and Δe which denote shifts between the parameter values for ρ and e used in the simulation of the at-sensor radiance and the parameter values used for normalization in Eq. (18). The simulation under a shift $\Delta\rho = 0.05$ can explain both the large variation between predicted t_{tot} in *noon* and *off-noon* the acquisitions. The shift and $\Delta e > 0$ furthermore explains the sloping behavior of the predicted transfer functions in all acquisitions (Fig. 15 (a)). Overall, this shows that SFMNN cannot fully disentangle the reflectance contribution from the atmospheric transfer. The regularization $\ell_{\mathbf{atm}}$ (Eq. (19)) proves to be insufficient in *off-noon* conditions. While we could not yet pinpoint the exact reasons for the difference in disentangling capacity of SFMNN in *noon* and *off-noon* conditions, we posit that there are three plausible reasons.

First, there is increasing model incompatibility under more diffuse conditions as could be seen in Fig. 5. The physical model used for signal reconstruction in Eq. (1) includes a diffuse contribution to the at-sensor signal, but the approximate PCA-based formulation of the atmospheric transfer lacks accuracy for the representation of τ_{do} and ρ_{so} which become more important under low SZA conditions.

Secondly, the worse performance may be due to data quality deterioration in lower light conditions. At lower SZA the diffuse contribution to the at-sensor radiance increases strongly even under clear-sky conditions when HyPlant data is normally recorded. An increasing diffuse component, however, reduces the signal-to-noise ratio of the at-sensor radiance as the pixelwise hyperspectral signal is spatially mixed.

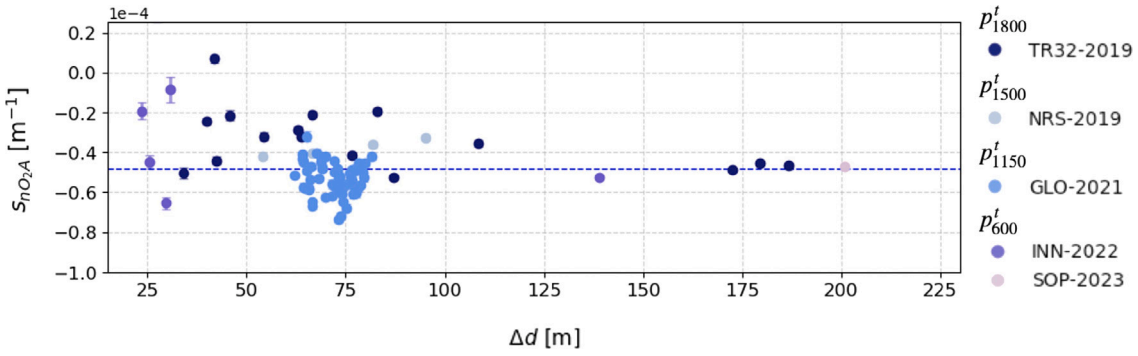


Fig. 13. Estimated s_{nO_2A} for all acquisitions with relevant variation in topography (cf. *topo* in Table 2). The dashed line represents the mean of five acquisitions with largest Δd. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

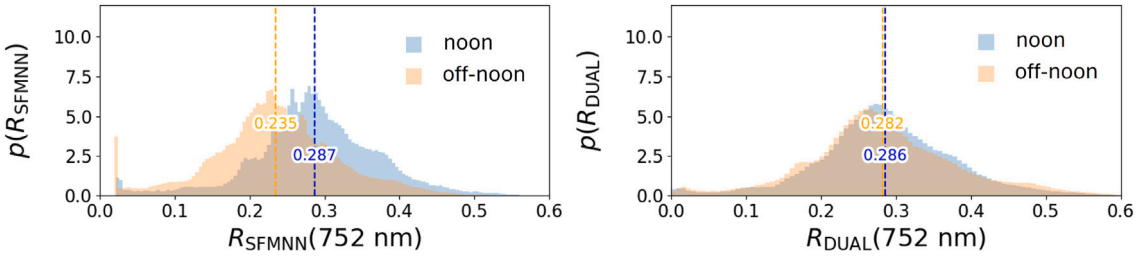


Fig. 14. Marginal reflectance distributions as predicted by SFMNN (a) and measured by DUAL (b) in the D_{day} data set. *Noon* and *off-noon* acquisition statistics are shown separately. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

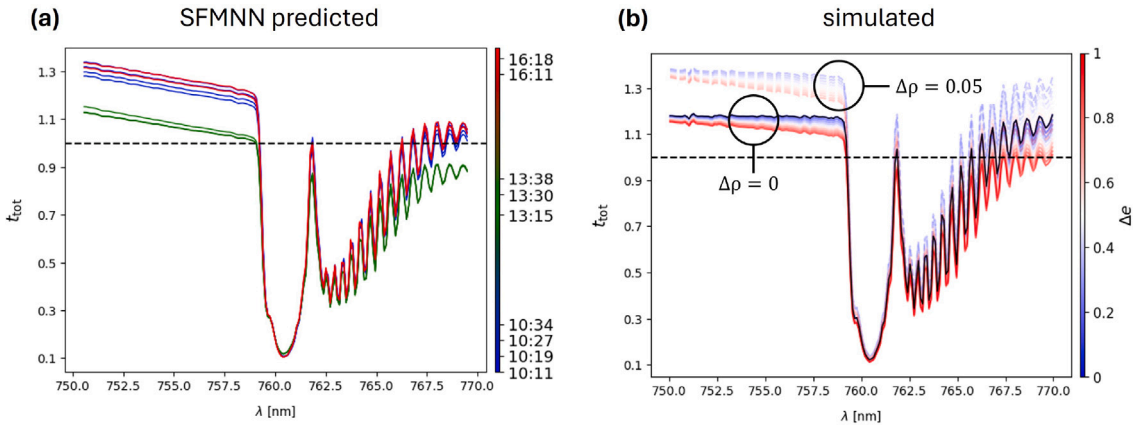


Fig. 15. **(a)** Mean SFMNN predicted atmospheric transfer t_{tot} in the nadir position in all acquisitions of D_{day}. **(b)** Simulated t_{tot} (black) simulation parameterization: AOT = 0.05, H_2O = 0.1, tilt angle TA = 0°, SZA = 30°, relative azimuth angle RAA = 30°, ground surface height h_{gnd} = 0.1 km, sensor height above ground h_{sen} = 1.6 km, ρ_{740} = 0.28, s_ρ = 0.007, e = 0. $\Delta\rho$ and Δe denote simulations where the normalizing reflectance in Eq. (18) was simulated with a shifted value with respect to the value used for at-sensor radiance calculation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The reduction in SNR is further enhanced due to the overall lower light intensity. This might reduce the ability of the network to disentangle reflectance and atmospheric components of the at-sensor signal.

Thirdly, the circumstance that the similitude of the predicted atmospheric transfer function with respect to our simulation in Fig. 15(b) is significantly improved under a change Δe indicates that the model does not represent surface reflectances accurately in the parameter space spanned by ρ, s_ρ and e (see Eq. (9)). This may in turn force the unphysical boosting of diffuse components.

The lack of control over the atmospheric decomposition is a significant draw-back of the PCA-based parameterization as compared to a physically based prediction of atmospheric components. This work

has only addressed the atmospheric at-sensor contribution in an aggregated way by analyzing the variation of the O_2-A band depth under topographic changes (Section 4.4) and the disentangling capacity of reflectance from atmospheric transfer in this section. Further work should therefore evaluate quantitatively the predicted atmospheric components and concentrate on proposing physical constraints regarding the disentangling of individual atmospheric components.

A promising possibility to constrain further the simulated atmospheric transfer consists in replacing the PCA-based forward simulation in SFMNN with a differentiable emulator of a physically accurate radiative transfer model. Research in biophysical modeling and simulation of hyperspectral reflectance from various platforms (Vicent et al., 2016; De Grave et al., 2020; Verrelst et al., 2021) are the basis for fast and accurate emulation models (Verrelst et al., 2016, 2017) and may be leveraged for computationally efficient retrieval of biophysical parameters. The emulation of at-sensor radiance simulations, as would be needed in our approach, has been addressed in Pato et al. (2024), Vicent et al. (2018). The differentiable nature of a number of those emulator models warrants an investigation into possibilities to implement such a replacement that would implement a tight integration of physically and statistically based SIF retrieval as has been shown recently for DESIS (Buffat et al., 2024a). Furthermore, the use of sensor-specific emulation facilitates the generalization of SFMNN to other airborne hyperspectral imaging sensors such as CFIS (Frankenberg et al., 2018), as well as to other spaceborne imaging sensors with appropriate spatial and spectral resolution such as FLEX (Drusch et al., 2017). With the use of emulators the radiative transfer physics and sensor characteristics affecting the at-sensor radiance can be modularized by domain experts and do not need to be modelled by the network as has been the case in this work where the atmospheric transfer was reconstructed from PCA weights on the fly. Such a procedure can exclude non-physical or implausible efficiently by imposing hard constraints on the at-sensor radiance model.

5.2. Accuracy of the SFMNN top-of-canopy SIF product

Direct validation with high-quality predictions from top-of-canopy measurements is the gold-standard for any SIF retrieval scheme if the spatial resolution of the sensor data is high enough to resolve the measurement footprint of the top-of-canopy measurements. The acquisition of field data and its synchronization with sensor acquisitions is, however, costly and usually results in small and spatially very constrained validation data sets. In this work we have gathered the majority of FLOX derived top-of-canopy SIF estimates acquired in tandem with HyPlant flights covering the period 2018–2022 to thoroughly validate SFMNN with high-quality data (see Table 6). While this study is the largest and most complete systematic comparison of HyPlant derived SIF estimates with top-of-canopy FLOX measurements, the validation data set has still a limited size and mainly covers crop canopies. Moreover, in addition to the uncertainties on the geolocalization and time synchronization, that were considered in this work, other uncertainty sources can impact the comparison of HyPlant derived SIF and FLOX derived SIF. Most importantly, we could not quantify uncertainties related to different viewing angles of in-situ and airborne radiance measurements (Liu et al., 2016).

In our validation study we could show that SFMNN prediction resulted in a reduction of the absolute error with respect to SFM and an improvement of correlation scores r with respect to iFLD. With SFMNN we found the best correlation scores r overall. SFMNN, however, was shown to still be impacted by biases leading to overestimation of SIF in absolute terms that lead to a similar performance as iFLD in terms of MAE. iFLD, on the other hand, clearly underperformed in terms of r as compared to SFM and SFMNN.

We found SFMNN MAE covering the range from 0.2–0.7 mW nm^{-1} sr^{-1} m^{-2} depending on the campaign data set used for validation. The FLEX mission requirement for SIF retrieval accuracy of 0.2 mW nm^{-1} sr^{-1} m^{-2} (Vicent et al., 2016) could not be consistently achieved for HyPlant data with the current set-up using one of the three SIF retrieval methods (iFLD, SFM, SFMNN) that were investigated in this study.

5.3. Learning of implicit constraints

We have shown in two application studies that SFMNN predictions are constrained implicitly. The observation of a constraint on the second order derivative β in diurnal SIF products suggests that the network is able to learn a first order physiological relationship between day time (i.e. SZA) and SIF emission magnitude. On the other hand, it could be shown that the atmospheric transfer function could be adapted in a way to represent the variation of the optical path in acquisitions solely based on radiance data. The fact that the feature based internal representations learned by SFMNN can be shown to be sensitive to basic physiological and physical constraints emphasizes that the proposed loss formulation allows the network to generalize the disentangling in typical HyPlant recording conditions beyond a pure functional optimization as it is done in SFM. This aspect of SFMNN can be considered to be an important advantage of SFMNN over the methodologies followed by SFM and iFLD as it allows in principle the use of trained models on data that was not included in the training. If a single model could be applied directly to new data prediction times could be significantly reduced as compared to the approach followed here (pretraining and fine-tuning) as well as compared to related spectral fitting methods such as SFM. Further validation should therefore especially concentrate on the generalization capacity of SFMNN SIF prediction for an operational context where fast prediction times are advantageous.

The feature representation of the hyperspectral signal as learned by the SFMNN encoder is based on a restricted number of HyPlant acquisitions. Recent scientific breakthroughs in the domain of neural network training for vision applications draw however on the self-supervised, application-agnostic learning of feature representations from very large data sets. Recently, first adaptations of large Vision Transformer and Diffusion models have been developed for remote sensing RGB (Wang et al., 2023; Khanna et al., 2023) and multispectral imagery (Blumenstiel et al., 2024). The multitude of spaceborne, globally distributed hyperspectral datasets that have become and will be available in the near future (PRISMA Pignatti et al., 2013, EnMAP Guanter et al., 2015b, DESIS Krutz et al., 2019, FLEX Drusch et al., 2017, CHIME Celesti et al., 2022) are likely to allow similar training set-ups. A natural extension of the present work will therefore consist in adapting the training of the SFMNN encoder backbone to a large collection of hyperspectral data sets from different sensors and platforms to improve its representative power.

6. Conclusion

We have presented in this work SFMNN, a self-supervised deep learning method to estimate SIF in the O_2-A absorption band of hyperspectral HyPlant imagery. We have applied the method to HyPlant acquisitions from multiple years and different observational conditions and compared the results to both an SFM and an iFLD SIF retrieval method for HyPlant data by performing a validation with ground based FLOX measurements. This is the first time that HyPlant SIF retrieval methods and products have been compared systematically in a validation with a FLOX in-situ data set spanning multiple years. In this validation study with in-situ measured SIF in flat terrain we could show that SFMNN yielded state-of-the-art SIF predictions in terms of its correlation score outperforming both iFLD and SFM. In terms of accuracy we found that SFMNN is impacted by an overestimating bias. Despite this bias SFMNN outperformed SFM in terms of accuracy.

In a second study on predicted diurnal SIF variation we found that the expected diurnal SIF dynamics in vegetated and non-vegetated areas were physiologically plausible. We have shown that SFMNN could

learn implicitly a first order physiological constraint regarding day time (i.e. SZA) and SIF emission magnitude.

We could furthermore observe that SFMNN could be applied to HyPlant data taken in regions with large topographic variation. One of the most serious issues of the baselines considered in this work is to find a suitable parameterization of the atmospheric transfer in regions with changing atmospheric and observational conditions. As SFMNN can locally compensate such changes in principle, an increased interest was put in its ability to yield trustworthy SIF estimates in such conditions. Since no in-situ data was available for flight lines exhibiting large topographic changes, we could only conduct a qualitative analysis. We found that the effective extinction coefficient determined from the linear relationship between O_2-A band depth and the estimated radiative path length converged to a constant value in flight lines with large topographic variations.

CRediT authorship contribution statement

Jim Buffat: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Miguel Pato:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Kevin Alonso:** Writing – review & editing, Conceptualization. **Stefan Auer:** Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Emiliano Carmona:** Writing – review & editing, Conceptualization. **Stefan Maier:** Writing – review & editing, Conceptualization. **Rupert Müller:** Conceptualization. **Patrick Rademske:** Data curation. **Bastian Siegmann:** Writing – review & editing, Data curation. **Uwe Rascher:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization. **Hanno Scharr:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Jim Buffat reports financial support was provided by Helmholtz Artificial Intelligence Cooperation Unit. Miguel Pato reports financial support was provided by Helmholtz Artificial Intelligence Cooperation Unit. Kevin Alonso reports financial support was provided by Helmholtz Artificial Intelligence Cooperation Unit. Stefan Auer reports financial support was provided by Helmholtz Artificial Intelligence Cooperation Unit. Emiliano Carmona reports financial support was provided by Helmholtz Artificial Intelligence Cooperation Unit. Rupert Mueller reports financial support was provided by Helmholtz Artificial Intelligence Cooperation Unit. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is part of the project “FluoMap” (Impulsfonds-Förderkennzeichen ZT-I-PF-5-12) funded by the Helmholtz Initiative and Networking Fund, Helmholtz AI, Deutsches Zentrum für Luft- und Raumfahrt (DLR) and Forschungszentrum Jülich GmbH (FZJ). The authors gratefully acknowledge computing time on the supercomputer JURECA (Jülich Supercomputing Centre, 2021) at Forschungszentrum Jülich under grant no. fluomap-ct. We gratefully acknowledge the financial support by the European Space Agency (ESA) for airborne data acquisition and data analysis in the frame of the FLEXSense campaign (ESA Contract No. 4000125402/ 18/NL/NA) and the Photoproxy project (ESA contract No. 4000125731/19/NL/LF). Additionally, HyPlant and FLOX data acquisition has partially been funded by the German Federal Ministry of Education and Research within the German-Plant-Phenotyping Network (DPPN) (project identification number: 031A053), the ‘Strukturwandel-Projekt Bioökonomie REVIER’, which is funded by the German Federal Ministry of Education and Research (project identification number 031B0918A), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany’s Excellence Strategy – EXC 2070–390732324 and the project Land surface Interactions with the Atmosphere over the Iberian Semi-arid Environment (LIAISE) funded by the Centre national de la recherche scientifique (CNRS). Open access is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 491111487.

Appendix. Derivation of SNR-based loss weighting

Let us consider a simplified model for SIF estimation and assume we have N measurements y on two wavelengths λ_1 and λ_2. Signal is composed by a reflectance part $y_R(\lambda)$ and a fluorescence part $y_R(\lambda)$ including atmospheric influences, such that

$$y(\lambda) = y_R(\lambda) + y_f(\lambda) + \epsilon, \tag{A.1}$$

where ϵ denotes represents instrumental noise. We model $R(\lambda)$ and $f(\lambda)$ to be

$$R(\lambda_i) = c_R(\lambda_i) R_0 \tag{A.2}$$

$$f(\lambda_i) = c_f(\lambda_i) f_0 \tag{A.3}$$

with known, constant function $c_\cdot(\lambda)$ and scalar parameters R_0 and f_0. In a more realistic model there would be more than one parameter for R and f. Please note, that we can select $c_\cdot(\lambda)$ such that $\mathbb{E}[R_0] = 1$ and $\mathbb{E}[f_0] = 1$ if we want to. Our model equations for estimation are then

$$y(\lambda_1)_j = A_R(\lambda_1) c_R(\lambda_1) R_0 + A_f(\lambda_1) c_f(\lambda_1) f_0 + \epsilon_j \tag{A.4}$$

$$y(\lambda_2)_j = A_R(\lambda_2) c_R(\lambda_2) R_0 + A_f(\lambda_2) c_f(\lambda_2) f_0 + \epsilon_j \tag{A.5}$$

where A_R and A_f contain all atmospheric parameters and all the rest. Note, that the multiple measurements indicated by index j only change ϵ_j. Obviously, we can multiply each equation with a freely adaptable weight w_i without changing anything. We can rewrite the equation to be

$$\mathbf{W}\mathbf{y} = \mathbf{W}\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\epsilon} \tag{A.6}$$

where $\mathbf{W}$ is a diagonal weight matrix with $W_{ii} = w_i$, $X_{i1} = A_R(\lambda_i) c_R(\lambda_i)$, $X_{i2} = A_f(\lambda_i) c_f(\lambda_i)$, $\beta_1 = R_0$, and $\beta_2 = f_0$. Noise ϵ has $\mathbb{E}(\epsilon|\mathbf{X}) = 0$ and $\mathrm{cov}(\epsilon|\mathbf{X}) = \boldsymbol{\Omega}$. In our case we assume that noise is independent, such that $\boldsymbol{\Omega}$ is diagonal with $\Omega_{ii} = w_i^2 u_i^2$, where u_i is the noise standard deviation for wavelength i. Clearly, weights w_i also scale the noise in the weighted equations. We know from Gauss–Markov-Theorem that the generalized least-squares estimator can then be written as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{y} \tag{A.7}$$

where weights w_i cancel out. Using this simultaneous estimation of all parameters β_i for designing a suitable loss function is not obvious. We therefore investigate a special case, where only one parameter is unknown.

Let us assume R_0 to be known and we want to infer f_0. We can then reformulate (A.4) as

$$\tilde{\mathbf{y}} = \mathbf{X} f_0 + \boldsymbol{\epsilon} \tag{A.8}$$

where $\tilde{y}_i = y_i - y_{R,i}$ and $X_i = A_f(\lambda_i) c_f(\lambda_i)$. The BLUE is given by

$$\hat{f}_0 = \frac{\sum_i X_i \tilde{y}_i / u_i^2}{\sum_i X_i^2 / u_i^2} \tag{A.9}$$

or, more suggestively written

$$\hat{f}_0 = \sum_i \frac{1}{u_i^2} \frac{X_i}{\sum_j X_j^2 / u_j^2} \tilde{y}_i \tag{A.10}$$

For the usual Moore–Penrose pseudo-inverse we would get

$$\hat{f}_0 = \sum_i \frac{X_i}{\sum_j X_j^2} \tilde{y}_i \tag{A.11}$$

The difference between these two estimation approaches is in the weights

$$w_i = \frac{1}{u_i^2} \frac{\sum_j X_j^2}{\sum_j X_j^2/u_j^2} \tag{A.12}$$

For training, we derive a mean weighting function $w_f = \mathbb{E}_p[w_i(L_{\text{at-s}}, f)]$ over a specific distribution p of simulated at-sensor radiances $L_{\text{at-s}}$ and top-of-canopy fluorescence. The simulations are performed in this case with a simple two-stream model, a fixed atmospheric transfer function and varying reflectance and fluorescence functions. Given the HyPlant-specific sensor variance model σ^2, this expectation can be written as

$$w_f(\lambda) = \left\langle \frac{1}{u_\lambda^2} \frac{\sum_{\lambda'} f_{\lambda'}^2}{\sum_{\lambda'} f_{\lambda'}^2/u_{\lambda'}^2} \right\rangle_{p(L_{\text{at-s}}, f)}. \tag{A.13}$$

Data availability

The training and validation data will be made available. The DOI will be provided at a later stage of the review process.

References

Alonso, L., Gomez-Chova, L., Vila-Frances, J., Amoros-Lopez, J., Guanter, L., Calpe, J., Moreno, J., 2008. Improved fraunhofer line discrimination method for vegetation fluorescence quantification. IEEE Geosci. Remote. Sens. Lett. 5 (4), 620–624. http://dx.doi.org/10.1109/LGRS.2008.2001180.

Anderson, T.L., Charlson, R.J., Winker, D.M., Ogren, J.A., Holmén, K., 2003. Mesoscale variations of tropospheric aerosols*. J. Atmos. Sci. 60 (1), 119–136. http://dx.doi.org/10.1175/1520-0469(2003)060<0119:MVOTA>2.0.CO;2.

Bandopadhyay, S., Rastogi, A., Juszczak, R., 2020. Review of top-of-canopy sun-induced fluorescence (SIF) studies from ground, UAV, airborne to spaceborne observations. Sensors 20 (4), 1144. http://dx.doi.org/10.3390/s20041144.

Bandopadhyay, S., Rastogi, A., Rascher, U., Rademske, P., Schickling, A., Cogliati, S., Julitta, T., Mac Arthur, A., Hueni, A., Tomelleri, E., Celesti, M., Burkart, A., Stróżecki, M., Sakowska, K., Gąbka, M., Rosadziński, S., Sojka, M., Iordache, M.-D., Reusen, I., Van Der Tol, C., Damm, A., Schuettemeyer, D., Juszczak, R., 2019. Hyplant-derived sun-induced fluorescence—A new opportunity to disentangle complex vegetation signals from diverse vegetation types. Remote Sens. 11 (14), 1691. http://dx.doi.org/10.3390/rs11141691.

Bendig, J., Chang, C.Y., Wang, N., Atherton, J., Malenovský, Z., Rascher, U., 2021. Measuring solar-induced fluorescence from unmanned aircraft systems for operational use in plant phenotyping and precision farming. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. pp. 1921–1924. http://dx.doi.org/10.1109/IGARSS47720.2021.9555157.

Blumenstiel, B., Moor, V., Kienzler, R., Brunschwiler, T., 2024. Multi-spectral remote sensing image retrieval using geospatial foundation models. http://dx.doi.org/10.48550/arXiv.2403.02059, arXiv:2403.02059.

Buffat, J., Pato, M., Alonso, K., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H., 2023. Deep learning based prediction of sun-induced fluorescence from hyplant imagery. In: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium. IEEE, Pasadena, CA, USA, pp. 2993–2996. http://dx.doi.org/10.1109/IGARSS52108.2023.10282828.

Buffat, J., Pato, M., Alonso, K., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H., 2024a. Retrieval of sun-induced plant fluorescence in the o_2-A absorption band from DESIS imagery. arXiv:2411.08925, URL https://arxiv.org/abs/2411.08925.

Buffat, J., Rascher, U., Rademske, P., Siegmann, B., Junker-Frohn, L.V., Emin, D., 2024b. Hydata: HyPlant FLUO at-sensor radiance data packages and FLOX measurements for SIF retrieval method development from selected campaigns of the years 2018 - 2023. http://dx.doi.org/10.26165/JUELICH-DATA/QKJKPW.

Celesti, M., Rast, M., Adams, J., Boccia, V., Gascon, F., Isola, C., Nieke, J., 2022. The copernicus hyperspectral imaging mission for the environment (chime): status and planning. In: IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium. pp. 5011–5014. http://dx.doi.org/10.1109/IGARSS46834.2022.9883592.

Chang, C.Y., Guanter, L., Frankenberg, C., Köhler, P., Gu, L., Magney, T.S., Grossmann, K., Sun, Y., 2020. Systematic assessment of retrieval methods for canopy far-red solar-induced chlorophyll fluorescence using high-frequency automated field spectroscopy. J. Geophys. Res.: Biogeosci. 125 (7), http://dx.doi.org/10.1029/2019JG005533, e2019JG005533.

Chang, C.Y., Wen, J., Han, J., Kira, O., LeVonne, J., Melkonian, J., Riha, S.J., Skovira, J., Ng, S., Gu, L., Wood, J.D., Näthe, P., Sun, Y., 2021. Unpacking the drivers of diurnal dynamics of sun-induced chlorophyll fluorescence (SIF): canopy structure, plant physiology, instrument configuration and retrieval methods. Remote Sens. Environ. 265, 112672. http://dx.doi.org/10.1016/j.rse.2021.112672.

Chen, X., Huang, Y., Nie, C., Zhang, S., Wang, G., Chen, S., Chen, Z., 2022. A long-term reconstructed TROPOMI solar-induced fluorescence dataset using machine learning algorithms. Sci. Data 9 (1), 427. http://dx.doi.org/10.1038/s41597-022-01520-1.

Cheng, Y.-B., Middleton, E.M., Zhang, Q., Huemmrich, K.F., Campbell, P.K.E., Corp, L.A., Cook, B.D., Kustas, W.P., Daughtry, C.S., 2013. Integrating solar induced fluorescence and the photochemical reflectance index for estimating gross primary production in a cornfield. Remote Sens. 5 (12), 6857–6879. http://dx.doi.org/10.3390/rs5126857.

Cogliati, S., Celesti, M., Cesana, I., Miglietta, F., Genesio, L., Julitta, T., Schuettemeyer, D., Drusch, M., Rascher, U., Jurado, P., Colombo, R., 2019. A spectral fitting algorithm to retrieve the fluorescence spectrum from canopy radiance. Remote Sens. 11 (16), 1840. http://dx.doi.org/10.3390/rs11161840.

Cogliati, S., Colombo, R., Celesti, M., Tagliabue, G., Rascher, U., Schickling, A., Rademske, P., Alonso, L., Sabater, N., Schuettemeyer, D., Drusch, M., 2018. Red and far-red fluorescence emission retrieval from airborne high-resolution spectra collected by the hyplant-fluo sensor. In: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. pp. 3935–3938. http://dx.doi.org/10.1109/IGARSS.2018.8517758.

Cogliati, S., Verhoef, W., Kraft, S., Sabater, N., Alonso, L., Vicent, J., Moreno, J., Drusch, M., Colombo, R., 2015b. Retrieval of sun-induced fluorescence using advanced spectral fitting methods. Remote Sens. Environ. 169, 344–357. http://dx.doi.org/10.1016/j.rse.2015.08.022.

Colombo, R., Celesti, M., Bianchi, R., Campbell, P.K., Cogliati, S., Cook, B.D., Corp, L.A., Damm, A., Domec, J.C., Guanter, L., Julitta, T., Middleton, E.M., Noormets, A., Panigada, C., Pinto, F., Rascher, U., Rossini, M., Schickling, A., 2018. Variability of sun-induced chlorophyll fluorescence according to stand age-related processes in a managed loblolly pine forest. Global Change Biol. http://dx.doi.org/10.1111/gcb.14097.

Damm, A., Cogliati, S., Colombo, R., Fritsche, L., Genangeli, A., Genesio, L., Hanus, J., Peressotti, A., Rademske, P., Rascher, U., Schuettemeyer, D., Siegmann, B., Sturm, J., Miglietta, F., 2022. Response times of remote sensing measured sun-induced chlorophyll fluorescence, surface temperature and vegetation indices to evolving soil water limitation in a crop canopy. Remote Sens. Environ. 273, 112957. http://dx.doi.org/10.1016/j.rse.2022.112957.

Damm, A., Guanter, L., Laurent, V.C.E., Schaepman, M.E., Schickling, A., Rascher, U., 2014. FLD-based retrieval of sun-induced chlorophyll fluorescence from medium spectral resolution airborne spectroscopy data. Remote Sens. Environ. 147, 256–266. http://dx.doi.org/10.1016/j.rse.2014.03.009.

Damm, A., Guanter, L., Paul-Limoges, E., van der Tol, C., Hueni, A., Buchmann, N., Eugster, W., Ammann, C., Schaepman, M.E., 2015. Far-red sun-induced chlorophyll fluorescence shows ecosystem-specific relationships to gross primary production: An assessment based on observational and modeling approaches. Remote Sens. Environ. 166, 91–105. http://dx.doi.org/10.1016/j.rse.2015.06.004.

Damm, A., Paul-Limoges, E., Kükenbrink, D., Bachofen, C., Morsdorf, F., 2020. Remote sensing of forest gas exchange: Considerations derived from a tomographic perspective. Global Change Biol. 26 (4), http://dx.doi.org/10.1111/gcb.15007.

De Grave, C., Verrelst, J., Morcillo-Pallarés, P., Pipia, L., Rivera-Caicedo, J.P., Amin, E., Belda, S., Moreno, J., 2020. Quantifying vegetation biophysical variables from the sentinel-3/FLEX tandem mission: Evaluation of the synergy of OLCI and FLORIS data sources. Remote Sens. Environ. 251, 112101. http://dx.doi.org/10.1016/j.rse.2020.112101.

Drusch, M., Moreno, J., Del Bello, U., Franco, R., Goulas, Y., Huth, A., Kraft, S., Middleton, E.M., Miglietta, F., Mohammed, G., Nedbal, L., Rascher, U., Schüttemeyer, D., Verhoef, W., 2017. The fluorescence explorer mission concept—ESA's earth explorer 8. IEEE Trans. Geosci. Remote Sens. 55 (3), 1273–1284. http://dx.doi.org/10.1109/TGRS.2016.2621820.

European Space Agency, 2017a. Technical assistance for the deployment of an advanced hyperspectral imaging sensor during FLEX-EU. http://dx.doi.org/10.5270/ESA-20835d4.

European Space Agency, 2017b. Technical assistance for the deployment of an advanced hyperspectral imaging sensor during soyflex. http://dx.doi.org/10.5270/ESA-50a3dd4.

European Space Agency, 2018. Technical assistance for the deployment of an advanced hyperspectral imaging sensor during SoyFLEX2. http://dx.doi.org/10.5270/ESA-24b3118.

European Space Agency, 2019. Photoproxy: technical assistance for the photosynthetic-proxy experiment. http://dx.doi.org/10.57780/esa-bb0ea39.

Frankenberg, C., Butz, A., Toon, G.C., 2011. Disentangling chlorophyll fluorescence from atmospheric scattering effects in O2 A-band spectra of reflected sun-light. Geophys. Res. Lett. 38 (3), http://dx.doi.org/10.1029/2010GL045896.

Frankenberg, C., Köhler, P., Magney, T.S., Geier, S., Lawson, P., Schwochert, M., McDuffie, J., Drewry, D.T., Pavlick, R., Kuhnert, A., 2018. The chlorophyll fluorescence imaging spectrometer (CFIS), mapping far red fluorescence from aircraft. Remote Sens. Environ. 217, 523–536. http://dx.doi.org/10.1016/j.rse.2018.08.032.

Gentine, P., Alemohammad, S.H., 2018. Reconstructed solar-induced fluorescence: A machine learning vegetation product based on MODIS surface reflectance to reproduce GOME-2 solar-induced fluorescence. Geophys. Res. Lett. 45 (7), 3136–3146. http://dx.doi.org/10.1002/2017GL076294.

Grace, J., Nichol, C., Disney, M., Lewis, P., Quaife, T., Bowyer, P., 2007. Can we measure terrestrial photosynthesis from space directly, using spectral reflectance and fluorescence? Global Change Biol. 13 (7), 1484–1497. http://dx.doi.org/10.1111/j.1365-2486.2007.01352.x.

Guanter, L., Aben, I., Tol, P., Krijger, J.M., Hollstein, A., Köhler, P., Damm, A., Joiner, J., Frankenberg, C., Landgraf, J., 2015a. Potential of the tropospheric monitoring instrument (TROPOMI) onboard the sentinel-5 precursor for the monitoring of terrestrial chlorophyll fluorescence. Atmos. Meas. Tech. http://dx.doi.org/10.5194/amt-8-1337-2015.

Guanter, L., Alonso, L., Gómez-Chova, L., Meroni, M., Preusker, R., Fischer, J., Moreno, J., 2010. Developments for vegetation fluorescence retrieval from spaceborne high-resolution spectrometry in the O2-A and O2-B absorption bands. J. Geophys. Res.: Atmos. 115 (D19), http://dx.doi.org/10.1029/2009JD013716.

Guanter, L., Bacour, C., Schneider, A., Aben, I., Van Kempen, T.A., Maignan, F., Retscher, C., Köhler, P., Frankenberg, C., Joiner, J., Zhang, Y., 2021. The TROPOSIF global sun-induced fluorescence dataset from the sentinel-5P TROPOMI mission. Earth Syst. Sci. Data 13 (11), 5423–5440. http://dx.doi.org/10.5194/essd-13-5423-2021.

Guanter, L., Frankenberg, C., Dudhia, A., Lewis, P.E., Gómez-Dans, J., Kuze, A., Suto, H., Grainger, R.G., 2012. Retrieval and global assessment of terrestrial chlorophyll fluorescence from GOSAT space measurements. Remote Sens. Environ. 121, 236–251. http://dx.doi.org/10.1016/j.rse.2012.02.006.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S., Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., Van der Linden, S., Leitão, P.J., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., Sang, B., 2015b. The EnMAP spaceborne imaging spectroscopy mission for earth observation. Remote Sens. 7 (7), 8830–8857. http://dx.doi.org/10.3390/rs70708830.

Guanter, L., Rossini, M., Colombo, R., Meroni, M., Frankenberg, C., Lee, J.-E., Joiner, J., 2013. Using field spectroscopy to assess the potential of statistical approaches for the retrieval of sun-induced chlorophyll fluorescence from ground and space. Remote Sens. Environ. 133, 52–61. http://dx.doi.org/10.1016/j.rse.2013.01.017.

Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J.A., Frankenberg, C., Huete, A.R., Zarco-Tejada, P., Lee, J.-E., Moran, M.S., Ponce-Campos, G., Beer, C., Camps-Valls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker, J.M., Griffis, T.J., 2014. Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. Proc. Natl. Acad. Sci. 111 (14), E1327–E1333. http://dx.doi.org/10.1073/pnas.1320008111.

Ioffe, S., Szegedy, C., 2015. Batch normalization: accelerating deep network training by reducing internal covariate shift. http://dx.doi.org/10.48550/arXiv.1502.03167, arXiv:1502.03167.

Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A.P., Middleton, E.M., Huemmrich, K.F., Yoshida, Y., Frankenberg, C., 2013. Global monitoring of terrestrial chlorophyll fluorescence from moderate spectral resolution near-infrared satellite measurements: Methodology, simulations, and application to GOME-2. http://dx.doi.org/10.5194/amtd-6-3883-2013.

Joiner, J., Yoshida, Y., Guanter, L., Middleton, E.M., 2016. New methods for the retrieval of chlorophyll red fluorescence from hyperspectral satellite instruments: Simulations andapplication to GOME-2 and SCIAMACHY. Atmos. Meas. Tech. 9 (8), 3939–3967. http://dx.doi.org/10.5194/amt-9-3939-2016.

Joiner, J., Yoshida, Y., Vasilkov, A.P., Yoshida, Y., Corp, L.A., Middleton, E.M., 2011. First observations of global and seasonal terrestrial chlorophyll fluorescence from space. Biogeosciences 8 (3), 637–651. http://dx.doi.org/10.5194/bg-8-637-2011.

Jülich Supercomputing Centre, 2021. JURECA: data centric and booster modules implementing the modular supercomputing architecture at jülich supercomputing centre. J. Large-Scale Res. Facil. 7 (A182), http://dx.doi.org/10.17815/jlsrf-7-182.

Khanna, S., Liu, P., Zhou, L., Meng, C., Rombach, R., Burke, M., Lobell, D., Ermon, S., 2023. DiffusionSat: A generative foundation model for satellite imagery. arXiv:2312.03606.

Kim, J., Ryu, Y., Dechant, B., Lee, H., Kim, H.S., Kornfeld, A., Berry, J.A., 2021. Solar-induced chlorophyll fluorescence is non-linearly related to canopy photosynthesis in a temperate evergreen needleleaf forest during the fall transition. Remote Sens. Environ. 258, 112362. http://dx.doi.org/10.1016/j.rse.2021.112362.

Kingma, D.P., Ba, J., 2017. Adam: A method for stochastic optimization. http://dx.doi.org/10.48550/arXiv.1412.6980, arXiv:1412.6980.

Krämer, J., Siegmann, B., Kraska, T., Muller, O., Rascher, U., 2021. The potential of spatial aggregation to extract remotely sensed sun-induced fluorescence (SIF) of small-sized experimental plots for applications in crop phenotyping. Int. J. Appl. Earth Obs. Geoinf. 104, 102565. http://dx.doi.org/10.1016/j.jag.2021.102565.

Krutz, D., Müller, R., Knodt, U., Günther, B., Walter, I., Sebastian, I., Säuberlich, T., Reulke, R., Carmona, E., Eckardt, A., Venus, H., Fischer, C., Zender, B., Arloth, S., Lieder, M., Neidhardt, M., Grote, U., Schrandt, F., Gelmi, S., Wojtkowiak, A., 2019. The instrument design of the DLR earth sensing imaging spectrometer (DESIS). Sensors 19 (7), 1622. http://dx.doi.org/10.3390/s19071622.

Kurucz, R.L., 2006. High resolution irradiance spectrum from 300 to 1000 nm. http://dx.doi.org/10.48550/ARXIV.ASTRO-PH/0605029, URL https://arxiv.org/abs/astro-ph/0605029.

Lee, J.E., Berry, J.A., van der Tol, C., Yang, X., Guanter, L., Damm, A., Baker, I., Frankenberg, C., 2015. Simulations of chlorophyll fluorescence incorporated into the community land model version 4. Global Change Biol. http://dx.doi.org/10.1111/gcb.12948.

Liu, L., Liu, X., Wang, Z., Zhang, B., 2016. Measurement and analysis of bidirectional SIF emissions in wheat canopies. IEEE Trans. Geosci. Remote Sens. 54 (5), 2640–2651. http://dx.doi.org/10.1109/TGRS.2015.2504089.

Maes, W.H., Pagán, B.R., Martens, B., Gentine, P., Guanter, L., Steppe, K., Verhoest, N.E.C., Dorigo, W., Li, X., Xiao, J., Miralles, D.G., 2020. Sun-induced fluorescence closely linked to ecosystem transpiration as evidenced by satellite data and radiative transfer models. Remote Sens. Environ. 249, 112030. http://dx.doi.org/10.1016/j.rse.2020.112030.

Magney, T.S., Frankenberg, C., Köhler, P., North, G., Davis, T.S., Dold, C., Dutta, D., Fisher, J.B., Grossmann, K., Harrington, A., Hatfield, J., Stutz, J., Sun, Y., Porcar-Castell, A., 2019. Disentangling changes in the spectral shape of chlorophyll fluorescence: implications for remote sensing of photosynthesis. J. Geophys. Res.: Biogeosci. 124 (6), 1491–1507. http://dx.doi.org/10.1029/2019JG005029.

Maier, S.W., Günther, K.P., Stellmes, M., 2004. Sun-induced fluorescence: a new tool for precision farming. In: Digital Imaging and Spectral Techniques: Applications To Precision Agriculture and Crop Physiology. John Wiley & Sons, Ltd, pp. 207–222. http://dx.doi.org/10.2134/asaspecpub66.c16.

Mazzoni, M., Meroni, M., Fortunato, C., Colombo, R., Verhoef, W., 2012. Retrieval of maize canopy fluorescence and reflectance by spectral fitting in the O2–A absorption band. Remote Sens. Environ. 124, 72–82. http://dx.doi.org/10.1016/j.rse.2012.04.025.

Meroni, M., Busetto, L., Colombo, R., Guanter, L., Moreno, J., Verhoef, W., 2010. Performance of spectral fitting methods for vegetation fluorescence quantification. Remote Sens. Environ. 114 (2), 363–374. http://dx.doi.org/10.1016/j.rse.2009.09.010.

Mohammed, G.H., Colombo, R., Middleton, E.M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Pérez-Priego, O., Damm, A., Meroni, M., Joiner, J., Cogliati, S., Verhoef, W., Malenovský, Z., Gastellu-Etchegorry, J.-P., Miller, J.R., Guanter, L., Moreno, J., Moya, I., Berry, J.A., Frankenberg, C., Zarco-Tejada, P.J., 2019. Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. Remote Sens. Environ. 231, 111177. http://dx.doi.org/10.1016/j.rse.2019.04.030.

Naethe, P., De Sanctis, A., Burkart, A., Campbell, P.K., Colombo, R., Di Mauro, B., Damm, A., El-Madany, T., Fava, F., Gamon, J.A., Huemmrich, K.F., Migliavacca, M., Paul-Limoges, E., Rascher, U., Rossini, M., Schüttemeyer, D., Tagliabue, G., Zhang, Y., Julitta, T., 2024. Towards a standardized, ground-based network of hyperspectral measurements: Combining time series from autonomous field spectrometers with sentinel-2. Remote Sens. Environ. 303, 114013. http://dx.doi.org/10.1016/j.rse.2024.114013.

Naethe, P., Julitta, T., Chang, C.Y.-Y., Burkart, A., Migliavacca, M., Guanter, L., Rascher, U., 2022. A precise method unaffected by atmospheric reabsorption for ground-based retrieval of red and far-red sun-induced chlorophyll fluorescence. Agricult. Forest. Meterol. 325, 109152. http://dx.doi.org/10.1016/j.agrformet.2022.109152.

NASA JPL, 2013. NASA shuttle radar topography mission global 1 arc second. http://dx.doi.org/10.5067/MEASURES/SRTM/SRTMGL1.003.

Nordrhein-Westfalen, 2017. DTM 1m. Sonny's LiDAR digital terrain models of Europe, https://sonny.4lima.de/, Accessed: 2025/01/08.

Pato, M., Alonso, K., Auer, S., Buffat, J., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H., 2023. Fast machine learning simulator of at-sensor radiances for solar-induced fluorescence retrieval with DESIS and hyplant. In: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium. pp. 7563–7566. http://dx.doi.org/10.1109/IGARSS52108.2023.10281579.

Pato, M., Buffat, J., Alonso, K., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H., 2024. Physics-based machine learning emulator of At-sensor radiances for solar-induced fluorescence retrieval in the O_2-A absorption band. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 1–10. http://dx.doi.org/10.1109/JSTARS.2024.3457231.

Pierrat, Z., Magney, T., Parazoo, N.C., Grossmann, K., Bowling, D.R., Seibt, U., Johnson, B., Helgason, W., Barr, A., Bortnik, J., Norton, A., Maguire, A., Frankenberg, C., Stutz, J., 2022. Diurnal and seasonal dynamics of solar-induced chlorophyll fluorescence, vegetation indices, and gross primary productivity in the boreal forest. J. Geophys. Res.: Biogeosci. 127 (2), http://dx.doi.org/10.1029/2021JG006588, e2021JG006588.

Pignatti, S., Palombo, A., Pascucci, S., Romano, F., Santini, F., Simoniello, T., Umberto, A., Vincenzo, C., Acito, N., Diani, M., Matteoli, S., Corsini, G., Casa, R., De Bonis, R., Laneve, G., Ananasso, C., 2013. The PRISMA hyperspectral mission: science activities and opportunities for agriculture and land monitoring. In: 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS. pp. 4558–4561. http://dx.doi.org/10.1109/IGARSS.2013.6723850.

Porcar-Castell, A., Tyystjärvi, E., Atherton, J., van der Tol, C., Flexas, J., Pfündel, E.E., Moreno, J., Frankenberg, C., Berry, J.A., 2014. Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: Mechanisms and challenges. J. Exp. Bot. 65 (15), 4065–4095. http://dx.doi.org/10.1093/jxb/eru191.

Rascher, U., Agati, G., Alonso, L., Cecchi, G., Champagne, S., Colombo, R., Damm, A., Daumard, F., de Miguel, E., Fernandez, G., Franch, B., Franke, J., Gerbig, C., Gioli, B., Gómez, J.A., Goulas, Y., Guanter, L., Gutiérrez-de-la-Cámara, Ó., Hamdi, K., Hostert, P., Jiménez, M., Kosvancova, M., Lognoli, D., Meroni, M., Miglietta, F., Moersch, A., Moreno, J., Moya, I., Neininger, B., Okujeni, A., Ounis, A., Palombi, L., Raimondi, V., Schickling, A., Sobrino, J.A., Stellmes, M., Toci, G., Toscano, P., Udelhoven, T., van der Linden, S., Zaldei, A., 2009. CEFLES2: The remote sensing component to quantify photosynthetic efficiency from the leaf to the region by measuring sun-induced fluorescence in the oxygen absorption bands. Biogeosciences 6 (7), 1181–1198. http://dx.doi.org/10.5194/bg-6-1181-2009.

Rascher, U., Baum, S., Siegmann, B., Buffat, J., Burkart, A., Cogliati, S., Colombo, R., Damm, A., Genesio, L., Hanus, J., Herrera, D., Julitta, T., Knopf, O., Miglietta, F., Muller, O., Quiros, J., 2022a. Flexsense: technical assistance for airborne measurements during the FLEX sentinel tandem experiment. http://dx.doi.org/10.57780/esa-84e5bf5.

Rascher, U., Siegmann, B., Baum, S., 2022b. Hyplant FLEX simulator on SAFIRE ATR42 for LIAISE experiment.

Rascher, U., Siegmann, B., Krieger, V., Matveeva, M., Quiros, J., Muller, O., Rademske, P., Herrera, D., Baum, S., Miglietta, F., Genesio, L., Colombo, R., Celesti, M., Tudoroiu, M., Cogliati, S., Carotenuto, F., Gioli, B., Genangelli, A., Hanus, J., Houlous, K., Migliavacca, M., Martini, D., Martin, M.P., Carrara, A., Moreno, G., Gonzales-Gascon, R., El-Madany, T., Pacheco-Labrador, J., Damm, A., Ahmed, R., Gupana, R., Paul-Limoges, E., Reiter, I., Xueref-Remy, I., Mevy, J.-P., Santonja, M., Bendig, J., Malenkovsky, Z., Julitta, T., Burkart, A., 2021. FLEX sentinel tandem campaign: technical assistance for airborne measurements during the FLEX sentinel tandem experiment. http://dx.doi.org/10.57780/esa-ae7953d.

Rossini, M., Nedbal, L., Guanter, L., Ač, A., Alonso, L., Burkart, A., Cogliati, S., Colombo, R., Damm, A., Drusch, M., Hanus, J., Janoutova, R., Julitta, T., Kokkalis, P., Moreno, J., Novotny, J., Panigada, C., Pinto, F., Schickling, A., Schüttemeyer, D., Zemek, F., Rascher, U., 2015. Red and far red sun-induced chlorophyll fluorescence as a measure of plant photosynthesis. Geophys. Res. Lett. 42 (6), 1632–1639. http://dx.doi.org/10.1002/2014GL062943.

Ryu, Y., Berry, J.A., Baldocchi, D.D., 2019. What is global photosynthesis? history, uncertainties and opportunities. Remote Sens. Environ. 223, 95–114. http://dx.doi.org/10.1016/j.rse.2019.01.016.

Sabater, N., Kolmonen, P., Van Wittenberghe, S., Arola, A., Moreno, J., 2021. Challenges in the atmospheric characterization for the retrieval of spectrally resolved fluorescence and PRI region dynamics from space. Remote Sens. Environ. 254, 112226. http://dx.doi.org/10.1016/j.rse.2020.112226.

Siegmann, B., Alonso, L., Celesti, M., Cogliati, S., Colombo, R., Damm, A., Douglas, S., Guanter, L., Hanuš, J., Kataja, K., Kraska, T., Matveeva, M., Moreno, J., Muller, O., Pikl, M., Pinto, F., Quirós Vargas, J., Rademske, P., Rodriguez-Morene, F., Sabater, N., Schickling, A., Schüttemeyer, D., Zemek, F., Rascher, U., 2019. The high-performance airborne imaging spectrometer hyplant—from raw images to top-of-canopy reflectance and fluorescence products: introduction of an automatized processing chain. Remote Sens. 11 (23), 2760. http://dx.doi.org/10.3390/rs11232760.

Siegmann, B., Cendrero-Mateo, M.P., Cogliati, S., Damm, A., Gamon, J., Herrera, D., Jedmowski, C., Junker-Frohn, L.V., Kraska, T., Muller, O., Rademske, P., van der Tol, C., Quiros-Vargas, J., Yang, P., Rascher, U., 2021. Downscaling of far-red solar-induced chlorophyll fluorescence of different crops from canopy to leaf level using a diurnal data set acquired by the airborne imaging spectrometer hyplant. Remote Sens. Environ. 264, 112609. http://dx.doi.org/10.1016/j.rse.2021.112609.

Subhash, N., Mohanan, C.N., 1997. Curve-fit analysis of chlorophyll fluorescence spectra: application to nutrient stress detection in sunflower. Remote Sens. Environ. 60 (3), 347–356. http://dx.doi.org/10.1016/S0034-4257(96)00217-9.

Sun, Y., Frankenberg, C., Jung, M., Joiner, J., Guanter, L., Köhler, P., Magney, T., 2018. Overview of solar-induced chlorophyll fluorescence (SIF) from the orbiting carbon observatory-2: retrieval, cross-mission comparison, and global monitoring for GPP. Remote Sens. Environ. http://dx.doi.org/10.1016/j.rse.2018.02.016.

Sun, Y., Frankenberg, C., Wood, J.D., Schimel, D.S., Jung, M., Guanter, L., Drewry, D.T., Verma, M., Porcar-Castell, A., Griffis, T.J., Gu, L., Magney, T.S., Köhler, P., Evans, B., Yuen, K., 2017. OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. Science http://dx.doi.org/10.1126/science.aam5747.

Tagliabue, G., Panigada, C., Dechant, B., Baret, F., Cogliati, S., Colombo, R., Migliavacca, M., Rademske, P., Schickling, A., Schüttemeyer, D., Verrelst, J., Rascher, U., Ryu, Y., Rossini, M., 2019. Exploring the spatial relationship between airborne-derived red and far-red sun-induced fluorescence and process-based GPP estimates in a forest ecosystem. Remote Sens. Environ. 231, 111272. http://dx.doi.org/10.1016/j.rse.2019.111272.

Tarquini, S., Isola, I., Favalli, M., Battistini, A., Dotta, G., 2023. TINITALY, a digital elevation model of Italy with a 10 meters cell size (version 1.1). http://dx.doi.org/10.13127/TINITALY/1.1.

Thompson, D.R., Bohn, N., Brodrick, P.G., Carmon, N., Eastwood, M.L., Eckert, R., Fichot, C.G., Harringmeyer, J.P., Nguyen, H.M., Simard, M., Thorpe, A.K., 2022. Atmospheric lengthscales for global VSWIR imaging spectroscopy. J. Geophys. Res.: Biogeosci. 127 (6), http://dx.doi.org/10.1029/2021JG006711, e2021JG006711.

Thompson, D.R., Kahn, B.H., Brodrick, P.G., Lebsock, M.D., Richardson, M., Green, R.O., 2021. Spectroscopic imaging of sub-kilometer spatial structure in lower-tropospheric water vapor. Atmos. Meas. Tech. 14 (4), 2827–2840. http://dx.doi.org/10.5194/amt-14-2827-2021.

van der Tol, C., Berry, J.A., Campbell, P.K.E., Rascher, U., 2014. Models of fluorescence and photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence. J. Geophys. Res.: Biogeosci. 119 (12), 2312–2327. http://dx.doi.org/10.1002/2014JG002713.

van der Tol, C., Julitta, T., Yang, P., Sabater, N., Reiter, I., Tudoroiu, M., Schuettemeyer, D., Drusch, M., 2023. Retrieval of chlorophyll fluorescence from a large distance using oxygen absorption bands. Remote Sens. Environ. 284, 113304. http://dx.doi.org/10.1016/j.rse.2022.113304.

van der Tol, C., Rossini, M., Cogliati, S., Verhoef, W., Colombo, R., Rascher, U., Mohammed, G., 2016. A model and measurement comparison of diurnal cycles of sun-induced chlorophyll fluorescence of crops. Remote Sens. Environ. 186, 663–677. http://dx.doi.org/10.1016/j.rse.2016.09.021.

van der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., Su, Z., 2009. An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. Biogeosciences 6 (12), 3109–3129. http://dx.doi.org/10.5194/bg-6-3109-2009.

Verrelst, J., De Grave, C., Amin, E., Reyes, P., Morata, M., Portales, E., Belda, S., Tagliabue, G., Panigada, C., Boschetti, M., Candiani, G., Segl, K., Guillasso, S., Berger, K., Wocher, M., Hank, T., Rascher, U., Isola, C., 2021. Prototyping vegetation traits models in the context of the hyperspectral chime mission preparation. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. pp. 7678–7681. http://dx.doi.org/10.1109/IGARSS47720.2021.9554407.

Verrelst, J., Rivera, J.P., van der Tol, C., Magnani, F., Mohammed, G., Moreno, J., 2015. Global sensitivity analysis of the SCOPE model: what drives simulated canopy-leaving sun-induced fluorescence? Remote Sens. Environ. 166, 8–21. http://dx.doi.org/10.1016/j.rse.2015.06.002.

Verrelst, J., Rivera Caicedo, J.P., Muñoz-Marí, J., Camps-Valls, G., Moreno, J., 2017. SCOPE-based emulators for fast generation of synthetic canopy reflectance and sun-induced fluorescence spectra. Remote Sens. 9 (9), 927. http://dx.doi.org/10.3390/rs9090927.

Verrelst, J., Sabater, N., Rivera, J.P., Muñoz-Marí, J., Vicent, J., Camps-Valls, G., Moreno, J., 2016. Emulation of leaf, canopy and atmosphere radiative transfer models for fast global sensitivity analysis. Remote Sens. 8 (8), 673. http://dx.doi.org/10.3390/rs8080673.

Vicent, J., Sabater, N., Tenjo, C., Acarreta, J.R., Manzano, M., Rivera, J.P., Jurado, P., Franco, R., Alonso, L., Verrelst, J., Moreno, J., 2016. FLEX end-to-end mission performance simulator. IEEE Trans. Geosci. Remote Sens. 54 (7), 4215–4223. http://dx.doi.org/10.1109/TGRS.2016.2538300.

Vicent, J., Verrelst, J., Rivera-Caicedo, J.P., Sabater, N., Muñoz-Marí, J., Camps-Valls, G., Moreno, J., 2018. Emulation as an accurate alternative to interpolation in sampling radiative transfer codes. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (12), 4918–4931. http://dx.doi.org/10.1109/JSTARS.2018.2875330.

Wang, N., Siegmann, B., Rascher, U., Clevers, J.G.P.W., Muller, O., Bartholomeus, H., Bendig, J., Masiliūnas, D., Pude, R., Kooistra, L., 2022. Comparison of a UAV- and an airborne-based system to acquire far-red sun-induced chlorophyll fluorescence measurements over structurally different crops. Agricult. Forest. Meterol. 323, 109081. http://dx.doi.org/10.1016/j.agrformet.2022.109081.

Wang, N., Suomalainen, J., Bartholomeus, H., Kooistra, L., Masiliūnas, D., Clevers, J.G.P.W., 2021. Diurnal variation of sun-induced chlorophyll fluorescence of agricultural crops observed from a point-based spectrometer on a UAV. Int. J. Appl. Earth Obs. Geoinf. 96, 102276. http://dx.doi.org/10.1016/j.jag.2020.102276.

Wang, D., Zhang, Q., Xu, Y., Zhang, J., Du, B., Tao, D., Zhang, L., 2023. Advancing plain vision transformer toward remote sensing foundation model. IEEE Trans. Geosci. Remote Sens. 61, 1–15. http://dx.doi.org/10.1109/TGRS.2022.3222818.

Wieneke, S., Ahrends, H., Damm, A., Pinto, F., Stadler, A., Rossini, M., Rascher, U., 2016. Airborne based spectroscopy of red and far-red sun-induced chlorophyll fluorescence: Implications for improved estimates of gross primary productivity. Remote Sens. Environ. 184, 654–667. http://dx.doi.org/10.1016/j.rse.2016.07.025.

Wu, Y., He, K., 2018. Group normalization. (arXiv:1803.08494), http://dx.doi.org/10.48550/arXiv.1803.08494.

Yao, L., Yang, D., Liu, Y., Wang, J., Liu, L., Du, S., Cai, Z., Lu, N., Lyu, D., Wang, M., Yin, Z., Zheng, Y., 2021. A new global solar-induced chlorophyll fluorescence (SIF) data product from TanSat measurements. Adv. Atmos. Sci. 38 (3), 341–345. http://dx.doi.org/10.1007/s00376-020-0204-6.

Zarco-Tejada, P., González-Dugo, M., Fereres, E., 2016. Seasonal stability of chlorophyll fluorescence quantified from airborne hyperspectral imagery as an indicator of net photosynthesis in the context of precision agriculture. Remote Sens. Environ. 179, 89–103. http://dx.doi.org/10.1016/j.rse.2016.03.024.

Zhang, Y., Joiner, J., Alemohammad, S.H., Zhou, S., Gentine, P., 2018. A global spatially contiguous solar-induced fluorescence (CSIF) dataset using neural networks. Biogeosciences 15 (19), 5779–5800. http://dx.doi.org/10.5194/bg-15-5779-2018.

Appendix B

Publication II: Physics-based Machine Learning Emulator of At-sensor Radiances for Solar-induced Fluorescence Retrieval in the O_2-A Absorption Band

Accepted Manuscript and published online in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2024; DOI: 10.1109/JSTARS.2024.3457231

Miguel Pato[a], Jim Buffat[b], Kevin Alonso[c], Stefan Auer[a], Emiliano Carmona[a], Stefan Maier[a], Rupert Müller[a], Patrick Rademske[b], Uwe Rascher[b], Hanno Scharr[d]

[a]Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany
[b]Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Jülich, Germany
[c]Starion Group c/o European Space Agency (ESA), Largo Galileo Galilei, Frascati 00044, Italy
[d]Forschungszentrum Jülich GmbH, Institute of Advanced Simulations, IAS-8: Data Analytics and Machine Learning, Jülich, Germany

Corresponding author: Miguel Pato, Miguel.FigueiredoVazPato@dlr.de

Author contributions: **Miguel Pato**: Writing – original draft, Writing – review & editing, Supervision, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Project administration, Funding acquisition, Conceptualization. **Jim Buffat**: Writing – review & editing, Visualization, Validation, Methodology, Conceptualization. **Kevin Alonso**: Writing — review & editing, Conceptualization. **Stefan Auer**: Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Emiliano Carmona**: Writing – review & editing, Conceptualization. **Stefan Maier**: Writing – review & editing, Conceptualization. **Rupert Müller**: Conceptualization. **Patrick Rademske**: Data curation. **Uwe Rascher**: Writing – review & editing, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization. **Hanno Scharr**: Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Overall contribution by Jim Buffat: 20 %

Physics-Based Machine Learning Emulator of at-Sensor Radiances for Solar-Induced Fluorescence Retrieval in the O_2-A Absorption Band

Miguel Pato, Jim Buffat, Kevin Alonso, Stefan Auer, *Member, IEEE*, Emiliano Carmona, Stefan Maier, Rupert Müller, Patrick Rademske, Uwe Rascher, and Hanno Scharr

***Abstract*—The successful operation of airborne and space-based spectrometers in recent years holds the promise to map solar-induced fluorescence (SIF) accurately across the globe. Machine learning (ML) can play an important role in this effort, but its application to SIF retrieval methods is in part hindered by the need for time-consuming radiative transfer modeling to account for atmospheric effects. In this work, we address this difficulty and develop a fast and accurate physics-based ML emulator of at-sensor radiances around the O_2-A absorption band for the space-based DESIS and the airborne HyPlant spectrometers. Different ML models are trained on an extensive set of simulated spectra encompassing a wide range of atmosphere, geometry, surface, and sensor configurations. A fourth-degree polynomial model is found to perform best, presenting errors at or below 10% of typical SIF at-sensor radiances and a prediction time per sample spectrum of 10−20 μs. Using data acquired with the HyPlant instrument, the proposed model is also shown to be able to match very closely the measured spectra. We illustrate how to improve further the accuracy of the emulator and how to generalize it to other sensors using the particular case of ESA's FLEX space mission. Our findings suggest that physics-based emulators can be efficiently used for the development of ML-based SIF retrieval methods by generating large training datasets in short time and by enabling a fast simulation step for self-supervised retrieval schemes.**

***Index Terms*—Hyperspectral sensors, machine learning, radiative transfer, solar-induced fluorescence.**

I. Introduction

REMOTELY sensed optical data collected by airborne or space-based Earth observation instruments inevitably carry the imprint of the atmosphere. These so-called atmospheric effects depend on wavelength range of interest, spectral resolution, and observation conditions, but usually they are substantial and cannot be ignored when analyzing the spectrometric data. Therefore, the processing of remote sensing data requires complex radiative transfer modeling in order to account or correct for atmospheric effects before surface-related observables can be derived. This step typically relies in part on specialized radiative transfer packages (such as MODTRAN [1], [2], libRadtran [3], [4], or 6S/6SV [5], [6], [7]), which are crucial in accurately modeling the light path through the atmosphere but are often too slow to be included in online correction or retrieval schemes. In many applications, the time-consuming radiative transfer calculations are performed offline and saved in look-up tables that can be swiftly accessed during correction or retrieval.

One such application is the retrieval of solar-induced fluorescence (SIF) emitted by plants between 650 and 800 nm during the photosynthetic process [8], [9]. SIF represents an important indicator of the efficiency of photosynthesis and can be used to monitor vegetation status. Thus, there is considerable interest in measuring SIF from remote platforms in order to study extended areas of vegetation. The SIF radiance is however much smaller than the surface reflected radiance [8], making its measurement particularly challenging even under favorable conditions. The relative strength of the SIF signal is significantly improved in regions of the spectrum where the reflected signal is reduced as in solar Fraunhofer lines or in atmospheric absorption bands. To resolve these features requires in turn a high spectral resolution instrument with sufficient signal-to-noise ratio. Despite the challenges, SIF has been successfully retrieved from selected airborne instruments both in solar Fraunhofer lines and in oxygen absorption bands [10], [11], [12] (see also [13]). Retrieval schemes such as the spectral fitting method [14], [15] make use of precomputed look-up tables based on offline radiative transfer computations. Because the latter are slow, the look-up tables are necessarily restricted to a limited range of atmosphere and observation geometries. This limitation could be efficiently addressed by machine learning (ML) models since they are well suited to speed up the forward simulation step and thereby provide a fast alternative to full-fledged radiative transfer modeling (see, e.g., [16], [17], [18]). Besides extending the scope of traditional SIF retrieval methods, an ML emulator can also be seamlessly integrated into the recent efforts to develop ML-based SIF retrieval schemes [19], [20], [21], [22].

Received 29 April 2024; revised 2 August 2024; accepted 27 August 2024. Date of publication 10 September 2024; date of current version 18 October 2024. *(Corresponding author: Miguel Pato.)*

Miguel Pato, Stefan Auer, Emiliano Carmona, Stefan Maier, and Rupert Müller are with the Remote Sensing Technology Institute, German Aerospace Center, 82234 Oberpfaffenhofen, Germany (e-mail: Miguel.FigueiredoVazPato@dlr.de).

Jim Buffat, Patrick Rademske, and Uwe Rascher are with the Institute of Bio-and Geosciences, IBG-2: Plant Sciences, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany.

Kevin Alonso is with the Starion Group c/o European Space Agency, Largo Galileo Galilei, 00044 Frascati, Italy.

Hanno Scharr is with the Institute of Advanced Simulations, IAS-8: Data Analytics and Machine Learning, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany (e-mail: h.Scharr@fz-juelich.de).

Digital Object Identifier 10.1109/JSTARS.2024.3457231

The goal of the present work is to develop a physics-based ML emulator of at-sensor radiances in the vicinity of the O_2-A absorption band ($\lambda = 740 - 780$ nm) for use in SIF retrieval methods. The choice of the O_2-A band is motivated by the significant depth of the absorption feature and its wide use for SIF retrieval. The narrow range also helps reducing the complexity of the problem and enables the generation of large datasets of simulated spectra. In the future, the O_2-B absorption band (around 687 nm) and other spectral ranges should also be considered. We focus on two representative instruments currently in operation, namely the DESIS spectrometer [23], [24] operating on the International Space Station since 2018 and the airborne HyPlant spectrometer [12] which is the demonstrator for the FLEX space mission. The proposed emulator can also be generalized to other sensors in a straightforward manner.

The article is organized as follows. Section II introduces the simulated data and methodology adopted, while Section III reports the performance of all tested physics-based ML emulators and explores strategies to improve upon our proposed ML model. Section IV discusses the relevance of our results and the concluding remarks are given in Section V. Note that this manuscript is an extended and improved version of our previous work presented in [25]. In particular, we use here newly simulated datasets (see Section II-A), compare the emulator to data acquired by the HyPlant instrument (see Section IV), and implement emulator variants with improved accuracy performance (see Section III-C).

II. Data and Methods

A. Simulated Data

The first step in developing our physics-based ML emulator is to assemble a body of simulated at-sensor radiance spectra capturing a meaningful range of atmosphere conditions, observation geometries, surface properties and sensor characteristics. For our purposes, the at-sensor radiance signal has contributions due to light scattered by the atmosphere as well as light reflected and emitted by the surface [26]:

$$L_s = L_p + \frac{E_g^0 \rho T^{\uparrow}}{\pi(1-\rho S)} + L_F T^{\uparrow} \tag{1}$$

where L_p is the path radiance, E_g^0 is the global solar irradiance on the ground, $T^{\uparrow}$ is the total transmission coefficient from surface to sensor (comprising direct and diffuse components, $T^{\uparrow} = T_{\text{dir}}^{\uparrow} + T_{\text{dif}}^{\uparrow}$), and S is the spherical albedo of the atmosphere. The surface reflectance and SIF on-ground radiance are represented by ρ and L_F, respectively. The at-sensor radiance spectrum in sensor resolution is then obtained by convoluting the above expression with the appropriate spectral response functions. It is noteworthy that the atmospheric functions L_p, E_g^0, $T_{\text{dir}}^{\uparrow}$, $T_{\text{dif}}^{\uparrow}$, and S depend solely on atmosphere and geometry and can thus be computed separately and applied for different surface and sensor configurations. Accordingly, a two-module software tool was developed to simulate at-sensor radiance spectra. The atmosphere and geometry parameters are passed to the first module, which derives the atmospheric functions L_p, E_g^0, $T_{\text{dir}}^{\uparrow}$, $T_{\text{dif}}^{\uparrow}$, and S at very high spectral resolution. We modeled radiative transfer through the atmosphere with the line-by-line algorithm of MODTRAN6 [2], but other radiative transfer software can also be used in our generic approach. The surface and sensor properties are handled by the second module of our tool, which outputs the at-sensor radiance spectrum in sensor resolution for any given atmosphere, geometry, surface, and sensor configuration.

The developed simulation tool is generic and can be used for virtually any optical remote sensing application. In this work, the tool is employed for the simulation of DESIS and HyPlant at-sensor radiances between 740 and 780 nm in view of SIF retrieval applications in the O_2-A absorption band. The modeling of surface and sensor properties is tailored for our specific case. The surface reflectance spectrum is parameterized as a second-order polynomial, while the SIF emission spectrum is taken as a Gaussian of variable normalization F_{737} at $\lambda = 737$ nm and with 20 nm standard deviation. Using real data acquired by DESIS and HyPlant, these models were found to be realistic for different vegetation and soil land covers in the wavelength range around the O_2-A band, but are not intended for other applications or spectral regions. In addition, the DESIS and HyPlant sensor properties were carefully considered based on expert knowledge of their performance and calibration over recent years. With typical spectral sampling distances of 2.55 nm for DESIS and 0.11 nm for HyPlant, there are 13 DESIS and 349 HyPlant spectral bands in the range $\lambda = 740 - 780$ nm with full width at half maximum (FWHM) of approximately 3.5 nm for DESIS and 0.24 nm for HyPlant. For detailed instrument specifications, please refer to [12], [23], [24]. We characterize the pixels of each band by an average Gaussian spectral response function and allow for small additive changes of central wavelength (CW) and FWHM. In this way, the simulated spectra for DESIS and HyPlant are very realistic and closely follow the spectral performance of the instruments.

We aim at simulating at-sensor radiances precisely enough to be useful to retrieve the weak SIF signal. For typical SIF on-ground outputs of $F_{737} = 0.1 - 0.4$ mW/cm^2/sr/μm at $\lambda = 737$ nm, the corresponding SIF at-sensor radiance amounts to $0.01-0.4$ mW/cm^2/sr/μm in the range $\lambda = 740 - 780$ nm. This sets the sensitivity goal for our simulations. We therefore conducted a sensitivity analysis with the objective of pinning down all parameters of interest for our case. The resulting set of parameters define atmosphere (water vapor content H_2O, aerosol optical thickness at 550 nm AOT $_{550}$), geometry (tilt angle TA, sun zenith angle SZA, relative azimuth angle RAA, ground altitude h_{gnd}, sensor altitude h_{sen}), surface (reflectance at 740 nm ρ_{740}, reflectance slope at 740 nm s, ratio of reflectance slopes at 780 and 740 nm e) and sensor (CW shift δ_{CW}, FWHM change δ_{FWHM}) properties. The corresponding ranges are listed in Table I, where the cases of DESIS and HyPlant were treated separately given their operation specificities. For clarification, the tilt angle is the angle between the line-of-sight and the nadir direction and is essentially equal to the view zenith angle in

TABLE I
SPECIFICATION OF THE DESIS AND HYPLANT SIMULATED DATASETS

Specification		DESIS	HyPlant
Atmosphere	H_2O [cm]	0.3−5.0	0.3−3.0
	AOT_{550} []	0.02−0.30	0.02−0.30
Geometry	TA [°]	0−25	0−20
	SZA [°]	0−55	20−55
	RAA [°]	0−180	0−180
	h_{gnd} [m]	0−600	0−300
	h_{sen} [km]	100	0.659−0.691
Surface	ρ_{740} []	0.05−0.60	0.05−0.60
	s [nm^{-1}]	0−0.012	0−0.012
	e []	0−1	0−1
	F_{737}/F_0	0−0.8	0−0.8
Sensor	δ_{CW} [nm]	$[-1.75, +1.25]$	$[-0.080, +0.080]$
	δ_{FWHM} [nm]	$[-0.3, +0.3]$	$[-0.040, +0.040]$
Input dimensions		12	13
Number of bands		13	349
Number of samples		4.3×10^6	6.3×10^6
- uniform grid		1×10^6	3×10^6
- random		3×10^5	3×10^5
- Halton		3×10^6	3×10^6

our cases of interest. Note that in the table the sensor altitude is given with respect to the ground in the case of HyPlant and the SIF on-ground radiance is expressed in terms of $F_0 = 1$ mW/cm^2/sr/μm. Beyond the parameters specified in Table I, we adopted throughout the mid-latitude summer atmosphere, the rural aerosol model, an ozone content of 332 DU, and the TSIS-1 solar model [27].

Table I effectively defines the input parameter space for our simulated dataset, encompassing 12 dimensions for DESIS and 13 for HyPlant. It is crucial to uniformly cover these high-dimensional spaces in order to appropriately represent the influence of atmosphere, geometry, surface and sensor properties on the at-sensor radiance spectra. Uniform grid sampling, frequently used for the generation of look-up tables, is not efficient for more than a few dimensions leaving large chunks of the parameter space unexplored. The random and Halton [28], [29] methods provide much more uniform sampling in high dimensions with a limited number of samples (for other methods, see, e.g., [30]). Therefore, we sampled the input space with a combination of the uniform grid, random, and Halton methods and then ran the simulation tool to generate the at-sensor radiance spectra. The key specifications of the DESIS and HyPlant simulated datasets are detailed in Table I; a full account of the simulations used here is presented in [31]. For later comparison, the simulation of a single spectrum typically takes 1−4 min depending on the input configuration and this time is dominated by the radiative transfer part. The simulated datasets, generated using ten cores on a dedicated virtual machine (Intel(R) Xeon(R) Gold 6132 CPU @ 2.60 GHz, 64 GB RAM) and leveraging the decoupling of the two modules of the simulation tool, contain 4.3 million DESIS and 6.3 million HyPlant spectra. This body of simulated data constitutes the basis for training a physics-based ML emulator.

B. Methodology

The influence of atmosphere, geometry, surface, and sensor properties on the at-sensor radiance spectrum can be learned by a ML model trained on the simulated datasets presented in the previous section. In fact, this constitutes a regression problem in multiple dimensions:

$$L_s = F(x) \tag{2}$$

where the feature vector $x \in \mathbb{R}^p$ contains the parameters identified in Table I, the target vector $L_s \in \mathbb{R}^b$ is the at-sensor radiance spectrum in sensor resolution, and $F : \mathbb{R}^p \to \mathbb{R}^b$ is the mapping function to be learned ($F : \mathbb{R}^{12} \to \mathbb{R}^{13}$ for DESIS, $F : \mathbb{R}^{13} \to \mathbb{R}^{349}$ for HyPlant). Given the sheer number of input dimensions in our problem, simple models for F can capture a significant degree of complexity while being very fast at evaluation. With this observation in mind, we attempted to learn $F(x)$ with basic benchmark models, namely linear functions, polynomials, and shallow neural networks. For clarity and later reference, a polynomial of degree d is defined as

$$P_d(x) = \sum_{\mathbf{k} \in \mathcal{K}_d} w_{\mathbf{k}} \prod_{i=1}^{p} x_i^{k_i}$$

$$\mathcal{K}_d = \left\{ \mathbf{k} \in \mathbb{N}^p : \sum_{i=1}^{p} k_i \leq d \right\} \tag{3}$$

with $w_{\mathbf{k}} \in \mathbb{R}^b$ the polynomial coefficients. For the models above, we use least squares as loss function. The case of a linear function (polynomial of degree $d = 1$) is simply ordinary least squares (OLS). In addition, kernel ridge regression (KRR), Gaussian process regression (GPR), support vector regression, and k nearest neighbors were also applied to learn $F(x)$. No attempt was made to realize more complex ML models (including deep learning). This choice is justified by the very encouraging results obtained with the very simple ML models mentioned above (see Section III). A comparative analysis of the different models studied here with traditional look-up table interpolation methods is a relevant line of research that is left for future work. Note however that, unlike interpolation methods, the emulators studied here are very lightweight in terms of memory usage.

A fast and accurate emulator of at-sensor radiances needs to be complex enough to handle the range of simulated configurations and simple enough to be computationally efficient. It turns out that basic ML models are very effective in our application case of simulating DESIS and HyPlant data around the O_2-A absorption band. We trained and assessed in detail several benchmark ML models: OLS, second-, fourth- and fifth-order polynomials (P2, P4, P5) and neural networks with 128 × 64 (N2) and 128 × 64 × 32 (N3) hidden nodes, batch size 16, and rectified linear unit activations. The polynomial models were trained with ridge regression, while for the neural network stochastic gradient descent was used. KRR and GPR were also considered since they have shown good performance for the emulation task [17], [18], [20]. ML models based on support vector regression and k nearest neighbors (k=1,5,10) led to poor results in terms of

TABLE II
TIME AND ACCURACY PERFORMANCE OF DIFFERENT ML MODELS TRAINED ON THE DESIS AND HYPLANT SIMULATED DATA

Performance parameter	DESIS							
	OLS	P2	P4	P5	N2	N3	KRR	GPR
Test set MAE [mW/cm^2/sr/μm]	0.77	0.17	0.011	0.0032	0.040	0.037	0.043	0.039
Total training time	1.6 s	19 s	1.8 min	1.2 min	1.2 h	1.2 h	1.1 min	2.7 s
Prediction time per sample	0.06 μs	1.1 μs	11 μs	28 μs	56 μs	45 μs	0.1 ms	0.2 ms

Performance parameter	HyPlant							
	OLS	P2	P4	P5	N2	N3	KRR	GPR
Test set MAE [mW/cm^2/sr/μm]	0.59	0.092	0.0027	0.0013	0.027	0.020	0.018	0.025
Total training time	1.7 min	45 s	1.3 min	1.5 min	2.0 h	2.2 h	6.3 s	13 s
Prediction time per sample	1.3 μs	2.2 μs	17 μs	47 μs	44 μs	43 μs	72 μs	0.7 ms

accuracy and speed early on in our analysis and were accordingly disregarded for in-depth evaluation. The performance of each of the benchmark models was explored by considering a limited set of the relevant hyperparameters (including L2 norm penalty for OLS, P2, P4, P5, N2, and N3; hidden nodes, batch size, training epochs for N2 and N3; kernel function for KRR) in the validation dataset. Since the figures did not change dramatically by fine-tuning the hyperparameters, in the following we opted to focus on the simplest representative configuration for each model. In particular, the models shown all have zero L2 norm penalty.

The simulated datasets specified in Table I were split into training, validation, and test sets considering the properties of the different sampling techniques used. Uniform grid samples populate the input space very sparsely but include the borders, while Halton and random samples lie uniformly across the space but not at its borders. Ultimately, we are interested in realizing a physics-based ML emulator with good performance in the bulk of the input parameter space, not necessarily on its borders. The borders are however important for the training phase. Therefore, we assigned all uniform grid and Halton samples to the training set and split the random samples into validation and test sets. The training, validation, and test sets contain 4×10^6 (93.0%), 2.1×10^5 (4.9%), and 9×10^4 (2.1%) samples, respectively, for DESIS and 6×10^6 (95.2%), 2.1×10^5 (3.3%), and 9×10^4 (1.4%) samples, respectively, for HyPlant. Note that the baseline emulators presented in the next section do not require the use of the full training set to deliver excellent accuracy performance (see Section III-B for details). This should be taken into account when evaluating the apparent imbalance of the training, validation, and test sets.

The performance of each ML model was evaluated in terms of accuracy using the at-sensor radiance mean absolute error (MAE) and in terms of computational time using the prediction time per sample. These performance metrics are to be compared to the at-sensor radiance corresponding to a typical SIF output of $F_{737} = 0.4$ mW/cm^2/sr/μm and the simulation time of 1−4 min per spectrum. The training, validation, and test phases for each model were conducted on the same standard virtual machine as mentioned in Section II-A using one core only. No effort was made to speed up the performance of the models by parallelizing the implementation.

III. RESULTS

A. Evaluation of Different Models

The test set accuracy, total training time, and prediction time per sample for all benchmark models are reported in Table II. Comparing the accuracy of the different models in terms of MAE with the typical SIF at-sensor radiances 0.04−0.4mW/cm^2/sr/μm at $\lambda = 740 - 780$ nm (corresponding to a SIF on-ground radiance of $F_{737} = 0.4$ mW/cm^2/sr/μm at $\lambda = 737$ nm), it is clear that linear or quadratic functions (OLS, P2) are not complex enough to approximate the mapping function F and are therefore not appropriate for our purposes. The accuracies attained by N2, N3, KRR, and GPR are significantly better, but still of the order of the small fluorescence signal we aim to characterize. The results shown for KRR correspond to a third-order polynomial kernel; similar but slightly worse results were obtained when using a radial basis function kernel. Note that limited effort was put into exploring all possibilities for neural networks, KRR, and GPR. It is well possible that better results can be obtained with these models by fine-tuning their configuration and/or hyperparameters. Fourth-order polynomials seem to be adequate to represent the interplay between simulated data and atmosphere, geometry, sensor, and surface parameters. In fact, with an MAE of 0.011 mW/cm^2/sr/μm for DESIS and 0.0027 mW/cm^2/sr/μm for HyPlant, the P4 model provides an average accuracy well below typical SIF at-sensor radiances. Higher order polynomials perform even better at the cost of prediction time as illustrated by the P5 results in Table II, but we decided to use P4 as our baseline model since it is faster and its performance is already excellent. Our results suggest in particular that in our specific case deep learning models are not required to realize a very accurate ML emulator.

The training phase is relatively fast for all models with total training times of seconds for OLS, P2, and GPR, minutes for P4, P5, and KRR and hours for N2 and N3. Note that due to memory limitations some models were trained on a subset of the training data. The neural networks were trained on CPU,

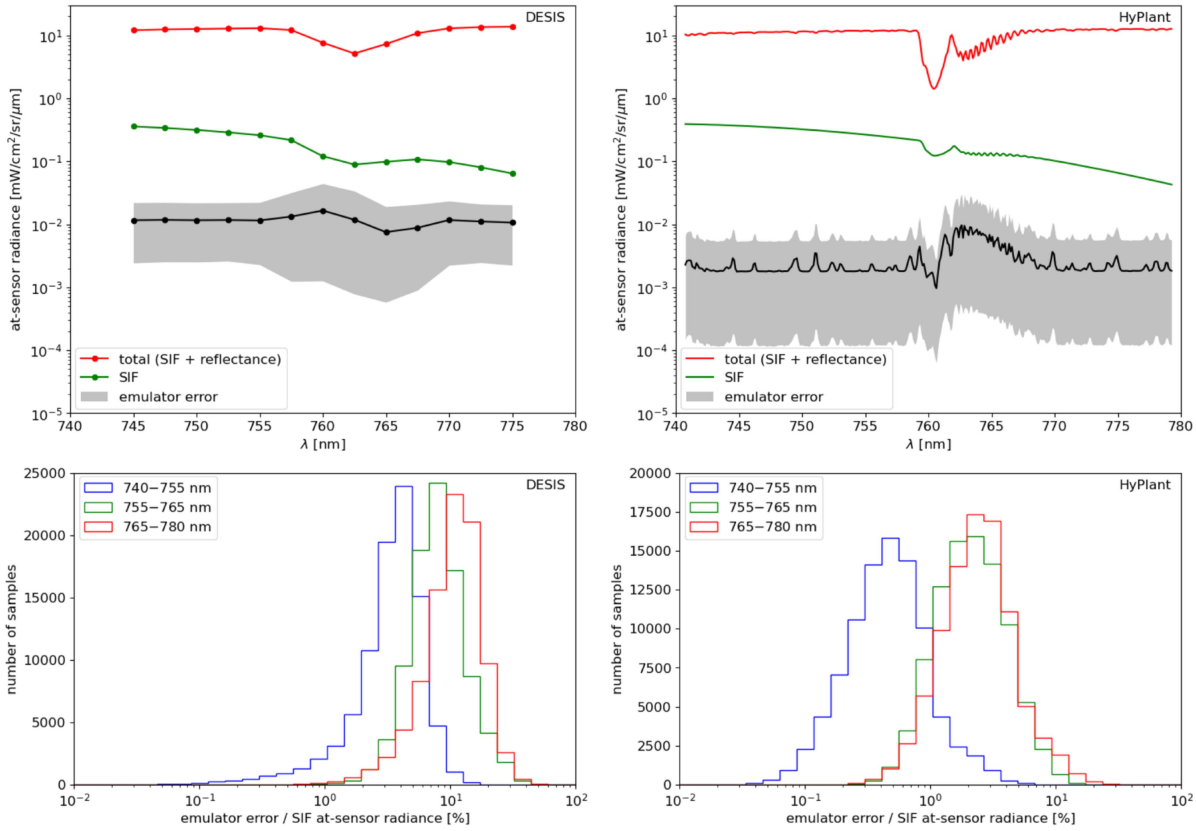


Fig. 1. Accuracy performance of the baseline P4 emulators trained on the DESIS (left) and HyPlant (right) simulated data. The top panels show the mean (black) and 5th/95th percentiles (gray) of the absolute error per band in the test set in direct comparison to the test set average total at-sensor radiance (red) and the SIF at-sensor radiance corresponding to $F_{737} = 0.4$ mW/cm^2/sr/μm (green). The distribution of the absolute error in the test set is plotted in the bottom panels relative to the SIF at-sensor radiance (top, green) for the spectral ranges around the O_2-A absorption band: $\lambda = 740 - 755$ nm (blue), $\lambda = 755 - 765$ nm (green), and $\lambda = 765 - 780$ nm (red).

so their time performance can likely be improved by the use of GPUs. Except for OLS, the times are similar for the cases of DESIS and HyPlant and do not scale with the number of output features (13 bands for DESIS, 349 for HyPlant). It is fair to point out that the total training time is not critical since training just occurs once; one can easily afford a longer training if a better accuracy can be achieved (see Section III-C for specific examples). More important is the prediction time per sample spectrum, also reported in Table II and computed as the mean time per spectrum when predicting on the training, validation, and test sets. OLS and P2 predictions take around 2 μs or less, while N2 and N3 prediction times are of order 50 μs (using CPU only). KRR and GPR prediction times are slightly higher at about 0.1 ms. But these models are not particularly useful for our purposes due to their reduced accuracy as shown above. The prediction time per spectrum for the proposed P4 model is of order $10-20$ μs. This is about 10^7 times faster than the simulation, which takes roughly $1-4$ min. The extreme speed up can be explained by the simplicity of the polynomial model used and the fact that prediction involves only matrix multiplications, which can be made very fast on a normal CPU. Note as well that for these models the predictions can be performed in bulk (e.g., for the whole test set). The sequential prediction of one spectrum at a time might differ from the values in Table II.

B. Baseline Emulator

It is clear from the discussion above that our best ML emulator is the fourth-order polynomial P4. The errors per band for this baseline model are shown in Fig. 1 separately for the DESIS and HyPlant cases. The upper plots compare the average total at-sensor radiances in the test set (red), the at-sensor radiance for a typical SIF signal with $F_{737} = 0.4$ mW/cm^2/sr/μm (green), and the P4 model error in the test set (black and gray) across the wavelength range $\lambda = 740 - 780$ nm. The mean test set error is represented by the black line, while the 5th and 95th percentiles are marked by the gray band and encompass the interval $5 \times 10^{-4} - 5 \times 10^{-2}$ mW/cm^2/sr/μm for DESIS and $10^{-4} - 10^{-2}$ mW/cm^2/sr/μm for HyPlant. Such error levels correspond to $\lesssim 0.1\%$ of the total at-sensor radiance, attesting the excellent accuracy of the P4 model. As evident from the bottom panels of Fig. 1, the model error is consistently at or below 10% of the typical SIF signal with $F_{737} = 0.4$ mW/cm^2/sr/μm for different wavelength ranges. Note in particular that there are virtually no test samples for which the error exceeds the typical SIF signal. The model performs similarly at and above the O_2-A band ($\lambda = 755 - 765$ nm and $\lambda = 765 - 780$ nm) and slightly better below the absorption band ($\lambda = 740 - 755$ nm). This is because the absolute errors are mostly constant across the

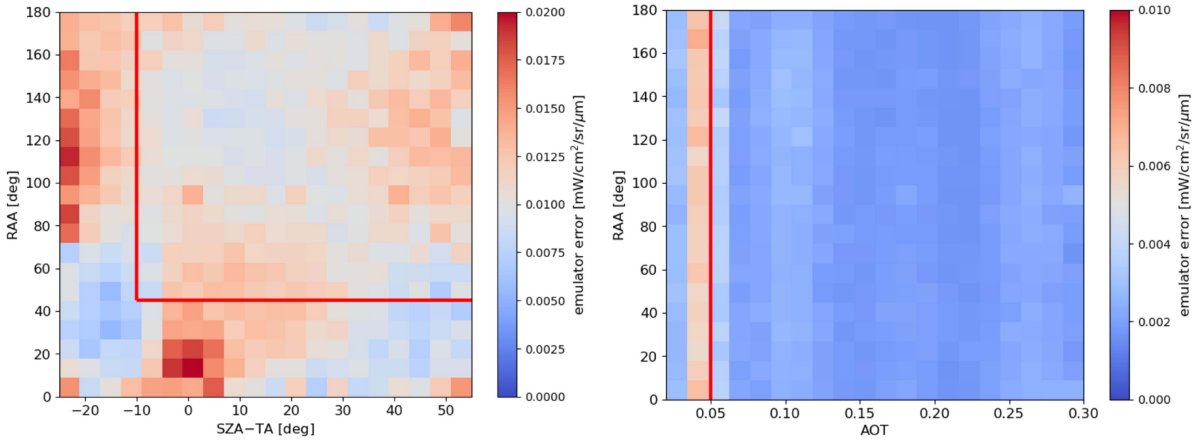


Fig. 2. Dependence of the accuracy of the baseline P4 emulator on observation conditions for DESIS (left) and HyPlant (right). The plots color-code the band-averaged absolute error in the training set as a function of SZA − TA and RAA for DESIS and of AOT_{550} and RAA for HyPlant. The regions to the left of the red lines encompass the largest errors and were masked out for the training and test of the P4A emulator variant.

spectral range considered (cf., upper panels of Fig. 1), while the SIF signal gradually decreases toward larger wavelengths. The main message conveyed by our results is that the P4 emulator is very accurate for the purposes of simulating the feeble SIF signal in the O_2-A band for DESIS and HyPlant.

For completeness, the performance of the baseline emulator is evaluated for different training set sizes. It turns out the P4 model is very accurate even when trained on small datasets. In fact, the accuracy of the emulator is the same or better than in Table II down to 5×10^3 training samples for DESIS and 5×10^4 training samples for HyPlant. For instance, using 5×10^4 randomly selected samples from the full training set, the test set MAE of the P4 model is 0.0061 mW/cm^2/sr/μm for DESIS (better than previously) and 0.0027 mW/cm^2/sr/μm for HyPlant (the same as previously). In this context, our test set with 9×10^4 samples, although a small fraction of the whole simulated dataset, is appropriate for testing the benchmark models. In addition, the results suggest that a significantly smaller training dataset would be sufficient for our particular application with the corresponding speed up of the simulation step and training of the emulator.

The performance of the P4 model is excellent across the whole input parameter space, but it is not uniform. In fact, there are parts of the parameter space where the emulator performs slightly worse than on average. This is illustrated in Fig. 2 for DESIS and HyPlant. In the case of DESIS, the most problematic region is related to sun backscatter geometries, where sun, sensor, and target are roughly aligned with the sun behind the sensor (i.e., SZA $\simeq$ TA and RAA $\simeq 0^\circ$). These configurations lead to an increase in sunlight backscattered by aerosols and are typically avoided if possible. For HyPlant, the problematic geometry does not occur since SZA $\geq$ TA (cf., Table I), but the largest residuals arise instead for a narrow range of small AOT $_{550}$ values. In both DESIS and HyPlant cases, the highest errors are very small (of order 0.02 mW/cm^2/sr/μm) and simply indicate where the selected model has a numerical difficulty in representing the simulated data.

C. Accuracy Improvement Strategies

Despite the encouraging findings presented above, it is important to understand the limitations of our proposed P4 emulator and explore ways to improve its performance. In the following, we consider variants of the P4 model with a restricted input space (P4A), bandwise learning (P4B), and atmospheric functions learning (P4C). The performance of the three variants is presented in Table III and Fig. 3.

We start by considering a restriction of the input space as a means to improve the performance of the P4 emulator. In fact, the nonuniform performance shown in Fig. 2 suggests that a better accuracy can be attained by partitioning the input parameter space and developing different models for each partition. In order to assess this possibility, we excluded the regions to the left of the red lines in Fig. 2, namely RAA $< 45^\circ$, SZA $-$ TA $< -10^\circ$ for DESIS and $AOT_{550} < 0.05$ for HyPlant, and retrained and tested a dedicated fourth-order polynomial model (P4A). As can be seen in Table III and Fig. 3, the P4A model improves upon the baseline P4 model on average by 21% for DESIS and 37% for HyPlant while being similarly fast. Although modest, the improvement is significant and illustrates how this low-effort strategy can boost the performance of a model in specific portions of the parameter space. Dedicated emulators for the masked problematic regions could also be developed if needed. Note that the trained emulators are only strictly valid in the corresponding input spaces and should not be used to interpret real data with parameters outside that space. This is especially important due to the polynomial nature of models which makes extrapolations beyond the input parameter space particularly prone to errors.

Next we explore the possibility to train our emulator bandwise. The formulation of the regression problem in Section II-B includes a multidimensional target (the at-sensor radiance spectrum) and implies the learning of all spectral bands in the same model. Alternatively, it is also possible to define a separate regression with a unidimensional target for each band. This

TABLE III
TIME AND ACCURACY PERFORMANCE OF VARIANTS OF THE BASELINE ML EMULATOR

Performance parameter	DESIS				HyPlant			
	P4	P4A	P4B	P4C	P4	P4A	P4B	P4C
Test set MAE [mW/cm^2/sr/μm]	0.011	0.0087	0.011	0.0037	0.0027	0.0017	0.0027	0.0014
Total training time	1.8 min	1.9 min	24 min	2.1 s	1.3 min	1.4 min	8.6 h	3 s
Prediction time per sample	11 μs	11 μs	0.1 ms	6 ms	17 μs	17 μs	4 ms	173 ms

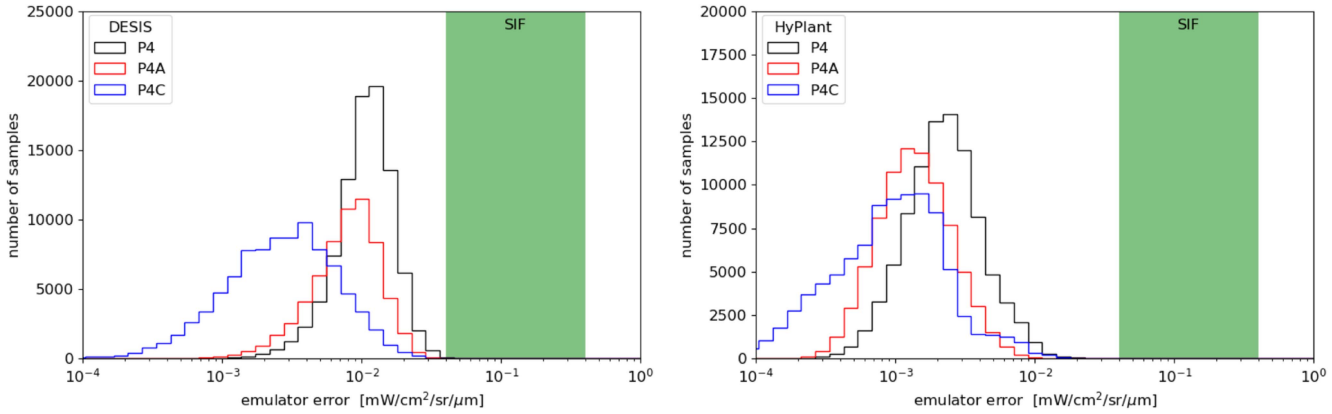


Fig. 3. Error distribution for the baseline P4 emulator and its variants for DESIS (left) and HyPlant (right). The histograms show the band-averaged absolute error in the test set for the baseline P4 model (black) and variants P4A (red, restricted input space) and P4C (blue, atmospheric functions learning). The results for the variant P4B (bandwise learning) are very similar to the ones of the baseline P4 model and are not shown for clarity. The green band indicates the SIF at-sensor radiance values corresponding to a typical SIF on-ground radiance of $F_{737} = 0.4$ mW/cm^2/sr/μm at $\lambda = 737$ nm.

approach decouples the different bands and enables the best possible fitting for each band at the price of additional computation time. The performance of a bandwise fourth-order polynomial model (P4B) is documented in Table III. As expected, training takes several times longer than for the baseline P4 model (especially for HyPlant given the large numbers of bands) and, more importantly, the prediction time per sample spectrum is 0.1 ms for DESIS and 4 ms for HyPlant, about 10 and 200 times slower than P4, respectively. Nevertheless, there is virtually no gain in accuracy of P4B with respect to P4 with MAE changes appearing only after the third significant digit. The results suggest that the overall regression with a multidimensional target used in P4 is very effective for both DESIS and HyPlant and it does not pay off to implement bandwise learning in our specific case.

Finally, we attempt to directly learn the atmospheric functions L_p, E_g^0, $T^\uparrow$, and S at very high spectral resolution instead of the sensor-resolution simulated data L_s [cf., (1)]. In other words, our regression problem is no longer as in (2) but it now reads

$$\left(L_p, E_g^0, T^\uparrow, S\right) = F(x)\,. \tag{4}$$

The crucial difference is that the new input space of $F(x)$ has a lower dimensionality. In fact, as explained in Section II-A, the atmospheric functions depend only on atmosphere and geometry parameters and can be used with any surface and sensor models for the derivation of the at-sensor radiance spectrum. Therefore, learning the atmospheric functions involves a regression function $F(x)$, where x contains only atmosphere and geometry parameters as opposed to the case of learning the simulated at-sensor radiance spectra in (2) where x contains all atmosphere, geometry, surface, and sensor parameters. This reduces significantly the dimensionality of the input parameter space: from 12 to 6 dimensions for DESIS and from 13 to 7 dimensions for HyPlant (cf., Table I). We trained a fourth-degree polynomial to learn the atmospheric functions (P4C) and combined it with a numerical module to factor in the surface and sensor properties as in (1). The performance of P4C model is detailed in Table III and Fig. 3. The training phase is now much quicker, because the dimensionality and size of the training set are much reduced with respect to the case of the P4 model. The prediction time is approximately 6 ms for DESIS and 173 ms for HyPlant, about three to four orders of magnitude slower than for P4. However, the average error in the test set is decreased by a factor of 3 to 0.0037 mW/cm^2/sr/μm for DESIS and by a factor of 2 to 0.0014 mW/cm^2/sr/μm for HyPlant. This suggests that the learning of atmospheric functions constitutes a promising strategy to realize a more accurate emulator at the price of a somewhat longer prediction. Whether a more accurate but slower ML-based emulator is useful for retrieval schemes depends on the specific application.

IV. DISCUSSION

In order to assess how well our emulators match observational data, we perform an unconstrained least-squares optimization

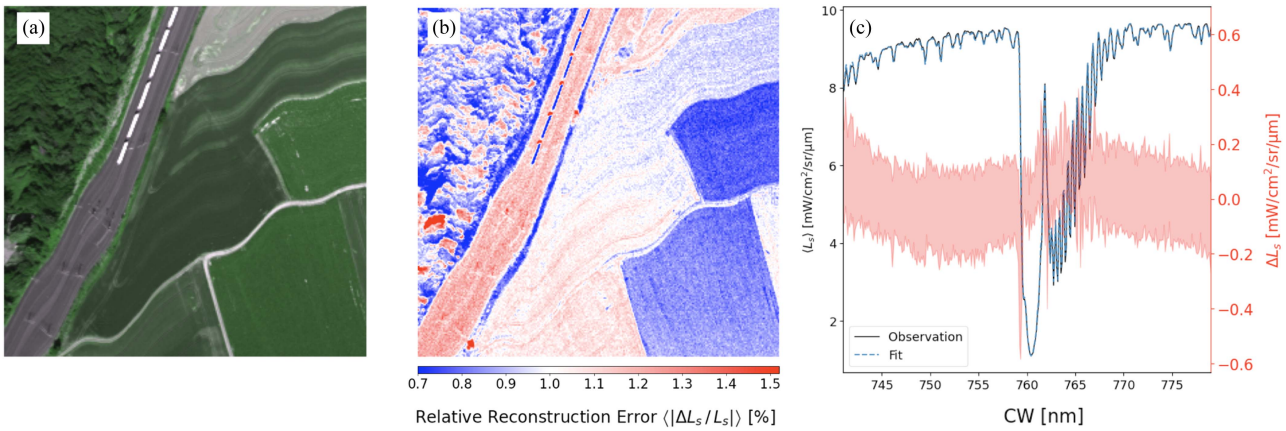


Fig. 4. Reconstruction of HyPlant spectra with the baseline P4 emulator model. The false color composite of a subset of a HyPlant acquisition (2018/06/26, 15:30 CEST, Jülich DE) is shown in (a), while the spectrally averaged relative reconstruction error is displayed in (b). Panel (c) plots the mean observed (black) and fitted (blue) at-sensor radiance within the subset as well as the 1th/99th percentiles of the reconstruction residuals (red).

of the input parameters of P4 on individual pixels of a HyPlant acquisition (cf., Fig. 4). The acquisition was recorded under optimal weather conditions such that its atmospheric state can be assumed to be covered by the simulation input ranges. Furthermore, we could infer geometry parameters from a digital elevation model and the sensor attitude recorded during acquisition which were covered by the simulation input ranges as well. Accordingly, the optimization was restricted to the atmosphere, surface, and sensor parameters (cf., Table I). We find the spectrally averaged reconstruction error to be $\langle|\Delta L_s|\rangle < 0.25$ mW/cm^2/sr/μm, which corresponds to a relative reconstruction error below 1.4%. It is clear from Fig. 4 that this performance holds across different surface classes commonly found in HyPlant acquisitions (e.g., crops, forested areas, man-made structures, and bare soil). Retrieval of physical parameters from spectral data is ill-posed due to the confounding nature of the parameters affecting the at-sensor radiance. As a consequence, the parameters derived as the optimal parametrization to P4 in each pixel cannot be trusted to approximate well the physical quantities underlying the at-sensor radiance signal generation. However, good reconstruction performance is a necessary precondition for emulators such as P4 to be used in various SIF retrieval approaches. Fig. 4 demonstrates the high accuracy of the baseline P4 emulator when compared to real data from HyPlant and therefore its usefulness for integration into future SIF retrieval methods.

We conclude our discussion by commenting on an additional use of the learned atmospheric functions in the P4C emulator. Since the atmospheric functions are independent of surface and sensor properties (cf., Section II-A), the P4C model effectively opens the possibility to generalize the simulation to any reflectance and fluorescence models and any putative sensor. This point is illustrated explicitly by simulating data with spectral sampling distances and spectral resolutions resembling the near future FLEX mission in the vicinity of the O_2-A absorption band. Loosely based on the specifications for the FLORIS instrument [32] and the learned P4C model, we show in Fig. 5 several simulated at-sensor radiance spectra between 740 and 780 nm for different observation conditions. It is important to emphasize that ours is not a high-fidelity simulation of the performance of the FLORIS instrument, which covers not only the O_2-A band but also the O_2-B band in high resolution and a considerably larger spectral range in lower resolution. In particular, the emulators developed in our work cover a limited spectral range ($\lambda = 740 - 780$ nm) and cannot be employed directly as emulators for the FLEX mission. However, the results in Fig. 5 serve to illustrate how our emulator can be used and extended for sensors other than HyPlant and DESIS. The ability to realistically simulate at-sensor radiance spectra will become crucial to interpret the data collected by FLEX and other missions and ultimately retrieve the underlying SIF spectrum.

V. Conclusion

Physics-based ML emulators have the potential to speed up radiative transfer modelling and thereby widen the reach of retrieval methods in varied remote sensing applications. Our work follows ongoing efforts in the community toward this direction and focuses on the case of SIF retrieval in the O_2-A absorption band. We deliver in particular a simple ML emulator of at-sensor radiances which is both fast and accurate enough to characterize the SIF signal in the DESIS and HyPlant spectrometers. Several strategies are analyzed to improve even more the accuracy of the emulator in exchange for lower speeds at prediction time. The results are not only relevant for DESIS and HyPlant, but are easily transferable to other sensors, including FLEX. A relevant line of research for future work is to extend our emulators to other spectral ranges, especially the O_2-B absorption band. This opens up the prospect of simulating large amounts of realistic spectrometer data quickly and integrating a fast simulation step into SIF retrieval methods. Both possibilities will be instrumental in bringing the power of machine learning to SIF studies over the coming years. In the wider picture, our work constitutes an illustrative example on the use of ML to speed up physics-based calculations in remote sensing applications.

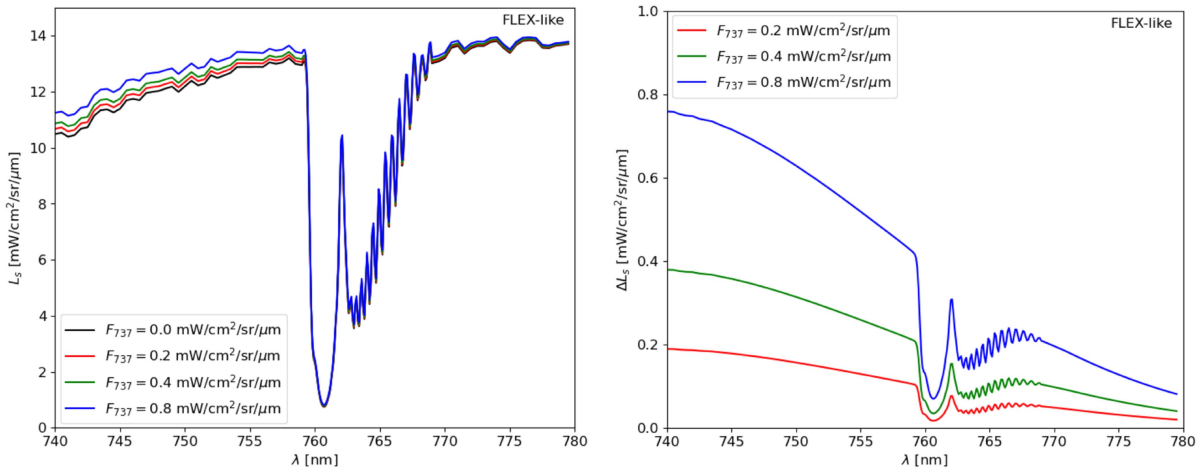


Fig. 5. Simulated at-sensor radiance spectra around the O_2-A absorption band for a FLEX-like instrument. The simulated spectra were obtained with the P4C emulator model for a space-based nadir observation with sun zenith angle of 30°, ground altitude of 0 m, aerosol optical thickness at 550 nm of 0.1, water vapor content of 2.0 cm, a slowly rolling reflectance spectrum ($\rho_{740} = 0.3$, $s = 0.006$ nm^{-1}, $e = 0$) and different SIF on-ground radiances ($F_{737} = 0.0, 0.2, 0.4, 0.8$ mW/cm^2/sr/μm). The spectral response function of the putative FLEX-like instrument was modelled roughly following the specifications of FLORIS [32] in the range $\lambda = 740 - 780$ nm. The left panel shows the at-sensor radiance spectra for different SIF radiances, while the right panel displays the at-sensor radiance differences with respect to the case without fluorescence.

ACKNOWLEDGMENT

This work was supported by the project "FluoMap" (Impulsfonds-Förderkennzeichen ZT-I-PF-5-12) funded by the Helmholtz Initiative and Networking Fund, Helmholtz AI, Deutsches Zentrum für Luft-und Raumfahrt and Forschungszentrum Jülich GmbH. The authors would like to acknowledge the computing time granted by the JARA Vergabegremium and provided on the JARA Partition part of the supercomputer JURECA [33] at Forschungszentrum Jülich.

REFERENCES

[1] A. Berk, P. Conforti, R. Kennett, T. Perkins, F. Hawes, and J. van den Bosch, "MODTRAN6: A major upgrade of the MODTRAN radiative transfer code," in *Proc. SPIE*, Jun. 2014, Art. no. 90880H.

[2] A. Berk, P. Conforti, and F. Hawes, "An accelerated line-by-line option for MODTRAN combining on-the-fly generation of line center absorption within 0.1 cm-1 bins and pre-computed line tails," in *Proc. SPIE*, May 2015, Art. no. 947217.

[3] C. Emde et al., "The libRadtran software package for radiative transfer calculations (version 2.0.1)," *Geosci. Model Develop.*, vol. 9, no. 5, pp. 1647–1672, 2016. [Online]. Available: https://gmd.copernicus.org/articles/9/1647/2016/

[4] B. Mayer and A. Kylling, "Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use," *Atmos. Chem. Phys.*, vol. 5, no. 7, pp. 1855–1877, 2005. [Online]. Available: https://acp.copernicus.org/articles/5/1855/2005/

[5] E. Vermote, D. Tanre, J. Deuze, M. Herman, and J.-J. Morcette, "Second simulation of the satellite signal in the solar spectrum, 6S: An overview," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 675–686, May 1997.

[6] S. Y. Kotchenova, E. F. Vermote, R. Matarrese, and F. J. Klemm, "Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path radiance," *Appl. Opt.*, vol. 45, no. 26, pp. 6762–6774, Sep. 2006. [Online]. Available: https://opg.optica.org/ao/abstract.cfm?URI=ao-45-26-6762

[7] S. Y. Kotchenova and E. F. Vermote, "Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part II. Homogeneous Lambertian and anisotropic surfaces," *Appl. Opt.*, vol. 46, no. 20, pp. 4455–4464, Jul. 2007. [Online]. Available: https://opg.optica.org/ao/abstract.cfm?URI=ao-46-20-4455

[8] G. H. Mohammed et al., "Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress," *Remote Sens. Environ.*, vol. 231, 2019, Art. no. 111177.

[9] A. Porcar-Castell et al., "Chlorophyll a fluorescence illuminates a path connecting plant molecular biology to earth-system science," *Nature Plants*, vol. 7, pp. 998–1009, 2021.

[10] U. Rascher et al., "Sun-induced fluorescence–A new probe of photosynthesis: First maps from the imaging spectrometer HyPlant," *Glob. Change Biol.*, vol. 21, no. 12, pp. 4673–4684, 2015.

[11] C. Frankenberg et al., "The chlorophyll fluorescence imaging spectrometer (CFIS), mapping far red fluorescence from aircraft," *Remote Sens. Environ.*, vol. 217, pp. 523–536, 2018.

[12] B. Siegmann et al., "The high-performance airborne imaging spectrometer hyplant-from raw images to top-of-canopy reflectance and fluorescence products: Introduction of an automatized processing chain," *Remote Sens.*, vol. 11, no. 23, 2019, Art. no. 2760.

[13] S. W. Maier, K. P. Günther, and M. Stellmes, *Sun-Induced Fluorescence: A New Tool for Precision Farming*. Madison, WI, USA: American Society of Agronomy, 2003, pp. 209–222.

[14] S. Cogliati et al., "Retrieval of sun-induced fluorescence using advanced spectral fitting methods," *Remote Sens. Environ.*, vol. 169, pp. 344–357, 2015.

[15] S. Cogliati et al., "A spectral fitting algorithm to retrieve the fluorescence spectrum from canopy radiance," *Remote Sens.*, vol. 11, no. 16, 2019, Art. no. 1840.

[16] J. V. Servera et al., "Gradient-based automatic lookup table generator for radiative transfer models," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1040–1048, Feb. 2019.

[17] J. Vicent Servera, L. Martino, J. Verrelst, and G. Camps-Valls, "Multifidelity Gaussian process emulation for atmospheric radiative transfer models," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5519210.

[18] J. Vicent Servera, L. Martino, J. Verrelst, J. P. Rivera-Caicedo, and G. Camps-Valls, "Multioutput feature selection for emulation and sensitivity analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5506411.

[19] R. Scodellaro et al., "A novel hybrid machine learning phasor-based approach to retrieve a full set of solar-induced fluorescence metrics and biophysical parameters," *Remote Sens. Environ.*, vol. 280, 2022, Art. no. 113196.

[20] J. Vicent Servera et al., "Systematic assessment of MODTRAN emulators for atmospheric correction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4101917.

[21] J. Buffat et al., "Deep learning based prediction of sun-induced fluorescence from hyplant imagery," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 2993–2996.

[22] J. Buffat et al., "A multi-layer perceptron approach for SIF retrieval in the O_2-A absorption band from hyperspectral imagery of the airborne hyplant sensor system," *Remote Sens. Environ.*, under review, 2024.
[23] D. Krutz et al., "The instrument design of the DLR earth sensing imaging spectrometer (DESIS)," *Sensors*, vol. 19, no. 7, 2019, Art. no. 1622. [Online]. Available: https://www.mdpi.com/1424-8220/19/7/1622
[24] K. Alonso et al., "Data products, quality and validation of the DLR earth sensing imaging spectrometer (DESIS)," *Sensors*, vol. 19, no. 20, 2019, Art. no. 4471.
[25] M. Pato et al., "Fast machine learning simulator of at-sensor radiances for solar-induced fluorescence retrieval with DESIS and hyplant," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2023, pp. 7563–7566.
[26] L. Guanter, R. Richter, and H. Kaufmann, "On the application of the MODTRAN4 atmospheric radiative transfer code to optical remote sensing," *Int. J. Remote Sens.*, vol. 30, no. 6, pp. 1407–1424, 2009.
[27] O. M. Coddington et al., "Version 2 of the TSIS-1 hybrid solar reference spectrum and extension to the full spectrum," *Earth Space Sci.*, vol. 10, no. 3, 2023, Art. no. e2022EA002637.
[28] J. Halton, "On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals," *Numerische Mathematik*, vol. 2, pp. 84–90, 1960.
[29] A. B. Owen, "A randomized Halton algorithm in R," 2017. [Online]. Available: https://arxiv.org/abs/1706.02808
[30] D. G. Loyola, R. M. Pedergnana, and S. Gimeno García, "Smart sampling and incremental function learning for very large high dimensional data," *Neural Netw.*, vol. 78, pp. 75–87, 2016.
[31] M. Pato et al., "Simulated at-sensor radiance data for solar-induced fluorescence retrieval with DESIS and HyPlant in the O2-A absorption band," *Remote Sens. Environ.*, to be submitted, 2024.
[32] M. Drusch et al., "The fluorescence explorer mission concept-ESA's earth explorer 8," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1273–1284, Mar. 2017.
[33] P. Thörnig, "JURECA: Data centric and booster modules implementing the modular supercomputing architecture at Jülich supercomputing centre," *J. Large-Scale Res. Facilities*, vol. 7, 2021, Art. no. A182.

Miguel Pato received the M.Sc. degree in physics engineering from the Technical University of Lisbon, Lisbon, Portugal, in 2007, and the joint Ph.D. degree in physics from the University of Padua, Padua, Italy, and Paris Diderot University, Paris, France, in 2011.

From 2011 to 2016, he was a Postdoctoral Researcher in the field of astroparticle physics with the University of Zurich, Zurich, Switzerland, Technical University of Munich, Munich, Germany, and Stockholm University, Stockholm, Sweden. Since 2017, he has been with the Department of Photogrammetry and Image Analysis, Remote Sensing Technology Institute, German Aerospace Center (DLR), Wessling, Germany. He is part of the EnMAP processing and calibration team. His research focuses on hyperspectral image analysis.

Jim Buffat received the M.Sc. degree in geography with specialization in remote sensing from the University of Zurich, Zurich, Switzerland, 2021. He is currently employed at Forschungszentrum Jülich, Jülich, Germany, working toward the Ph.D. in Engineering (Dr.-Ing.) awarded by the Faculty of Agriculture at the University of Bonn, Bonn, Germany.

His thesis centers on deep learning based retrieval of sun-induced plant fluorescence from hyperspectral imagery. His research interests include hyperspectral remote sensing, image processing, and deep learning for environmental applications.

Kevin Alonso received the Technical Telecommunication Engineering bachelor's degree in communication systems and the Telecommunication Engineering master's degree in telematic systems from the Bilbao School of Engineering, Bilbao, Spain, in 2005 and 2007, respectively, and the Ph.D. (Dr.-Ing.) degree, with thesis Heterogeneous Data Mining of Earth Observation Archives: Integration and Fusion of Images, Maps, and In-situ Data, from the Technical University of Munich (TUM), Munich, Germany, in 2017.

From July 2008 to March 2013, he was as a Staff Researcher with Vicomtech-IK4, where he worked in the field of digital television and multimedia services, participating in the European Research Project RUSHES. During 2009 and 2010, he combined his research work on semantic web and geographic information systems (GIS) technologies with the study of the master's degree in space science and technology. From 2013 to 2022, he was with the Department of Photogrammetry and Image Analysis, German Aerospace Center (DLR), Wessling, Germany. In September 2022, he joined Starion Group as a Mission Engineer to support the development of the Copernicus Expansion Missions CHIME and LSTM at ESA-ESRIN, Frascati, Italy.

Stefan Auer (Member, IEEE) received the Dipl.-Ing. degree in geodesy and the Ph.D. degree in remote sensing both from Technical University of Munich, Munich, Germany, in 2005 and 2011, respectively.

Since 2014, he has been a Senior Researcher and Project Manager with the Remote Sensing Technology Institute, German Aerospace Center, Wessling, Germany. Besides coordinating projects in the fields of data processing, image analysis, and machine learning, he pursues research in fusion techniques for multimodal remote sensing data, image interpretation, and scene modeling. He developed the open-source simulator RaySAR, which helps to understand the nature of prominent signatures in synthetic aperture radar images. In the context of his doctoral thesis, he spent three months as a guest Researcher with the Department of Electronic and Telecommunication Engineering, University of Naples "Federico II," Naples, Italy.

Emiliano Carmona received the M.Sc. degree in physics and the Ph.D. degree in astroparticle physics from the University of Valencia, Valencia, Spain, in 1997 and 2004, respectively.

From 2005 to 2013, he was a Postdoctoral Researcher with the Max-Planck-Institute for Physics, Munich, Germany, and a Researcher with CIEMAT-Centre for Energy, Environment and Technology, Madrid, Spain. Since 2014, he has been with the Department of Photogrammetry and Image Analysis, German Aerospace Center, Wessling, Germany. His research interests focus on hyperspectral remote sensing.

Stefan Maier received the B.Eng. and M.Eng. degrees in applied physics from the Scientific Technical Academy Isny, Allgau, Germany, in 1991 and 1994, respectively. He received the Doctorate in engineering from Technical University Munich, Munich, Germany, in 2000.

From 2000 to 2002, he was a Postdoctoral Researcher with the German Remote Sensing Data Center, German Aerospace Center DLR in Oberpfaffenhofen. From 2002 to 2007, he was a Senior Research Officer with Satellite Remote Sensing Services, Western Australian Land Information Authority, Perth, Australia. In 2007, he was a Manager of Satellite Remote Sensing Services, Western Australian Land Information Authority. From 2007 to 2015, he was Principal Research Fellow Remote Sensing with the Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, Australia. Since 2015, he has been the Director of maitec—Scientific and Engineering Consulting, Darwin, and since 2019 he has been a Visiting Scientist with the Department of Photogrammetry and Image Analysis, Remote Sensing Technology Institute, German Aerospace Center.

Rupert Müller received the Dipl.-Phys. degree in physics from the Ludwig Maximilians University of Munich, Munich, Germany, in 1984.

He is a senior scientist with the Remote Sensing Technology Institute, German Aerospace Center (DLR), Wessling, Germany, and Project Manager of the Ground Segment of MUSES (Multiuser System for Earth Sensing) with the hyperspectral instrument DESIS (DLR Earth Sensing Imaging Spectrometer) installed on the International Space Station. His main research interests include photogrammetric evaluation of optical data from air- and spaceborne sensors, digital image processing, machine learning, and hyperspectral imaging.

Patrick Rademske biography not available at the time of publication.

Uwe Rascher was born in Germany, in 1968. He received the Diploma degree in biology (master's) from University Erlangen-Nürnberg, Erlangen, Germany, in 1996, the Ph.D. degree (Dr.rer.nat.) in biology from the Technical University Darmstadt, Darmstadt, Germany, in 2001, and the Habilitation degree in plant biology from the University of Düsseldorf, Düsseldorf, Germany, in 2009.

From 2001 to 2004, he was working as a Postdoctoral Fellow with the Biosphere 2 Centre, Columbia University, Tucson, AZ, USA. In 2004, he joined the Forschungszentrum Jülich, Jülich, Germany, as a Young Research Scientist, where he was promoted to Group Leader in 2011. In 2014, he was appointed as a Full Professor with the Agricultural Faculty, University of Bonn, Bonn, Germany. He is currently the Head of the research area "Shoot Dynamics" at the Forschungszentrum Jülich, a Principal investigator of the Cluster of Excellence "PhenoRob," and a member of the MAG, European Space Agency (ESA), Darmstadt, Germany, in the preparation of the FLEX satellite mission, which will become the eighth Earth Explorer. His research combines work in plant physiology, remote sensing, ecology, global climate change, and plant phenotyping. His expertise is in characterizing photosynthesis using chlorophyll fluorescence, hyperspectral reflectance, gas exchange, remote sensing, and image analysis techniques. He further focuses on interdisciplinary work using novel remote sensing approaches to better measure photosynthesis and stress, develop novel plant phenotyping concepts, and model plant-mediated exchange processes from the leaf to the region.

Hanno Scharr received the Diploma and Ph.D. degrees in physics, from Heidelberg University, Heidelberg, Germany, in 1996 and 2000, respectively, and the Habilitation in computer science from Frankfurt University, Frankfurt, Germany, in 2014.

His Ph.D. work was on "Optimal Operators in Digital Image Processing." After a first postdoc in Heidelberg, he joined Intel Research, Santa Clara, CA, USA, for a year where he worked on real-time denoising algorithms for nanomachining tools. In 2003, he moved to Forschungszentrum, Jülich, Germany, leading a junior group on "Automatic Imaging and Quantitative Image Processing in Environment Plant Sciences" with a Young Investigator Award. He led the Image Analysis Group of the Institute of Bio- and Geosystems: IBG-2 Plant Sciences, Forschungszentrum Jülich. Since 2021, he has been the Head of the Institute of Advanced Simulation (IAS-8): Data Analytics and Machine Learning, Forschungzentrum Jülich. He has authored and coauthored more than 150 publications and patents. He serves as a Reviewer for all major computer vision journals, and for more than 35 international conference programme committes. His research interests include quantitative image processing, computer vision, and machine learning, with applications to natural sciences.

Appendix C

Publication III: Emulation-based self-supervised SIF retrieval in the O_2-A absorption band with HyPlant

Submitted Manuscript to Remote Sensing of Environment, 2025; Preprint published online, DOI: 10.22541/essoar.174000855.50541566/v2

Jim Buffat[a,d], Miguel Pato[b], Kevin Alonso[c], Stefan Auer[b], Emiliano Carmona[b], Stefan Maier[b], Rupert Müller[b], Patrick Rademske[a], Uwe Rascher[a], Hanno Scharr[d]

[a]Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Jülich, Germany
[b]Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany
[c]Starion Group c/o European Space Agency (ESA), Largo Galileo Galilei, Frascati 00044, Italy
[d]Forschungszentrum Jülich GmbH, Institute of Advanced Simulations, IAS-8: Data Analytics and Machine Learning, Jülich, Germany

Corresponding author: Jim Buffat, j.buffat@fz-juelich.de

Author contributions: **Jim Buffat**: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Miguel Pato**: Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Kevin Alonso**: Writing — review & editing, Conceptualization. **Stefan Auer**: Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Emiliano Carmona**: Writing – review & editing, Conceptualization. **Stefan Maier**: Writing – review & editing, Conceptualization. **Rupert Müller**: Conceptualization. **Patrick Rademske**: Data curation. **Uwe Rascher**: Writing – review & editing, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization. **Hanno Scharr**: Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Overall contribution by Jim Buffat: 90 %

Emulation-based self-supervised SIF retrieval in the O_2-A absorption band with HyPlant

Jim Buffat[a,d], Miguel Pato[b], Kevin Alonso[c], Stefan Auer[b], Emiliano Carmona[b], Stefan Maier[b], Rupert Müller[b], Patrick Rademske[d], Uwe Rascher[d], Hanno Scharr[a]

[a]*Forschungszentrum Jülich GmbH, Institute of Advanced Simulations, IAS-8: Data Analytics and Machine Learning, Jülich, Germany*
[b]*Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany*
[c]*European Space Agency (ESA), Largo Galileo Galilei, Frascati 00044, Italy*
[d]*Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Jülich, Germany*

Abstract

The retrieval of sun-induced fluorescence (SIF) from hyperspectral imagery requires accurate atmospheric compensation to correctly disentangle its small contribution to the at-sensor radiance from other confounding factors. In spectral fitting SIF retrieval approaches this compensation is estimated in a joint optimization of free variables when fitting the measured at-sensor signal. Due to the computational complexity of Radiative Transfer Models (RTMs) that satisfy the level of precision required for accurate SIF retrieval, fully joint estimations are practically inachievable with exact physical simulation. We present in this contribution an emulator-based spectral fitting method neural network (EmSFMNN) approach integrating RTM emulation and self-supervised training for computationally efficient and accurate SIF retrieval in the O_2-A absorption band of HyPlant imagery. In a validation study with in-situ top-of-canopy SIF measurements we find improved performance over traditional retrieval methods. Furthermore, we show that the model predicts plausible SIF emission in topographically variable terrain without scene-specific adaptations. Since EmSFMNN can be adapted to hyperspectral imaging sensors in a straightforward fashion, it may prove an interesting SIF retrieval method for other sensors on airborne and spaceborne platforms.

Keywords: Sun-induced fluorescence, deep learning, hyperspectral sensors, radiative transfer modelling, spectral fitting methods

1. Introduction

Any application based on hyperspectral imagery of the earth's surface acquired from remote platforms such as satellites, aircrafts or UAVs must consider the influence of the atmosphere at acquisition time. The atmospheric state has a confounding influence on the measured at-sensor radiance. In order to disentangle atmospheric effects from a physical surface variable of interest a firm understanding of the physical signal generation is necessary. Various atmospheric radiative transfer models (RTMs) have been developed (e.g. MODTRAN6 [1], 6S/6SV [2, 3], libRadTran [4]) to derive appropriate correction algorithms for these effects. In vegetation-related remote sensing it is crucial to couple such atmospheric models with leaf and soil optical properties, leaf-level energy fluxes [5, 6, 7, 8], and radiative transfer models in the canopy [9, 10] to enable accurate retrieval of biophysical parameters from remote sensing reflectances.

Similarly, the retrieval of sun-induced fluorescence (SIF) from hyperspectral imagery in atmospheric absorption bands relies heavily on accurate modelling of atmospheric radiative transfer and of sensor properties. The state of the atmosphere parameterized by its water vapour content, the type and density of aerosols at recording time as well as the pressure and temperature profiles along the optical path of the at-sensor signal modulate the radiance signal from which SIF is retrieved [11, 12, 13, 14]. Since in typical acquisition scenarios of hyperspectral at-sensor radiance for SIF retrieval no measurements are conducted to establish the signal contribution of the atmosphere during recording time, atmospheric variables must be estimated using RTMs in iterative processes. However, RTMs can often not be used directly in radiance-based estimation for individual pixels due to their computational cost. To reduce the retrieval dependency and the number of RTM simulations required to retrieve SIF from at-sensor radiance, a two-step procedure is assumed in various SIF retrieval methods as opposed to a joint estimation of surface, atmospheric and sensor related parameters. In a first step the atmosphere is characterized for a large number of pixels to derive the atmospheric transmittance with the help of an RTM. In a second step, these transmittance estimates are used to disentangle reflectance, fluorescence and possibly sensor miscalibrations commonly parametrized in center wavelength (CW) and full width at half maximum (FWHM) shifts. For example, [15, 16] derive a set of atmospheric transfer functions for single acquisitions using an RTM 'interrogation' technique first introduced by [17]. Operationally, these

estimated transfer functions are finetuned to account for retrieval errors of atmospheric components, sensor noise, sensor mischaracterization and model inaccuracies by modifications of a procedure called *transmittance correction* [18, 19]. This type of finetuning of the atmospheric transfer functions is based on the presence of non-vegetated pixels that are not affected by fluorescence. The identification of non-vegetated soil pixels can be difficult, however, in many geographical areas and especially in observation set-ups resulting in pixel sizes larger than a few meters where *pure* pixels are rare.

The validity of constant atmospheric transfer across a large set of spatial pixels relies on the fact that in airborne imagery the auto-correlation distance of atmospheric factors influencing the at-sensor radiance is usually larger than the spatial extent of the prediction [20, 21]. In the case of airborne acquisitions this results usually in the use of a single RTM estimate per acquisition. In the case of spaceborne acquisitions with a much larger spatial footprint, as will be provided for example by the FLEX mission [22], this assumption is not satisfied and strategies to localize the atmospheric characterizations efficiently must be developed. In the context of atmospheric correction for accurate reflectance estimation [23] have for example recently demonstrated the use of local linear emulators for accurate and computationally efficient atmospheric correction.

Similarly to changing atmospheric conditions on spatial scales relevant to satellite observations, the strongly changing observational conditions in airborne observations of topographically variable terrain are a challenge for SIF retrieval algorithms based on spectral regions affected by O_2 absorption. The simplifying assumption of constant atmospheric transmittance is invalid in these cases since the resulting optical path differences cause large variance in the depth of these absorption features.

[24] has proposed a pathway to computationally efficient SIF retrieval in these observational conditions. A reconstruction based on a Principal Component Analysis (PCA) of atmospheric transfer functions is used to model the radiative transfer non-parametrically. The use of PCA reconstructions allows for localized radiative transfer estimations and, importantly, a joint retrieval of the transfer functions as well as surface and sensor related quantities impacting the at-sensor radiance. However, the PCA loadings are fitted non-parametrically since they are not formulated as functions of physical quantities (e.g., surface and sensor altitude, water vapour content, aerosol optical density) as would be the case with physically explicit RTM simulations. This (i) impedes the explanatory power of atmospheric estimates and

(ii) does not allow for constraining the atmospheric estimates with known physical quantities.

In this work we propose the use of RTM emulation instead of the PCA-based radiative transfer formulation adopted in [24] to increase the physical accuracy of the Spectral Fitting Method Neural Network (SFMNN) in HyPlant FLUO data. Such an emulator-based SFMNN (EmSFMNN) approach has recently been validated for spaceborne DESIS data [25] in conjunction with a loss formulation similar to [24]. The authors find good agreement between their DESIS derived SIF product and HyPlant SIF estimates in a benchmark data set consisting of quasi-simultaneously recorded HyPlant and DESIS acquisitions highlighting the potential of this approach for hyperspectral sensors with improved spectral sampling intervals such as HyPlant FLUO and the FLORIS sensor onboard the ESA's Earth Explorer Mission FLEX [22].

RTM emulation can be regarded as a computationally efficient approximation of the exact RTM computation by a function acting on the same input parameter space as its RTM counterpart [26]. The functional form of such emulators is not relevant a priori, but depends on the specifications of the application such as the required computational speed and reconstruction performance, the spectral range of the application and the input parameter dimensionality. In this contribution, we derive a polynomial emulator from a large simulation database replicating typical observational conditions and the sensor characterization of the hyperspectral imaging sensor system HyPlant as in [11, 27, 28]. We additionally extend this emulator to represent bandwise spectral miscalibration which is shown to be integral for accurate SIF retrieval in HyPlant data. The functional form of this emulator matches well the specific requirements of neural network training. The computational efficiency of its predictions and gradient computation are sufficient for training on large hyperspectral data bases. With this novel neural network approach to integrate a computationally efficient model of canopy level optical properties and atmospheric radiative transfer into a SIF retrieval scheme we are able for the first time to make use of a pixelwise geometrical parameterization for a joint estimation of SIF and reflectance in airborne SIF retrieval.

In this study, We focus on SIF retrieval of selected campaign data sets of the hyperspectral HyPlant sensor system [29, 30]. The sensor characteristics of HyPlant and size of HyPlant data sets are uniquely suited to develop and improve partly data-driven SIF retrieval algorithms such as ours. Since HyPlant data is often acquired during field campaigns featuring ground based

Data Set	Campaign	FLOX	Δh [m]	$\lvert\mathcal{D}\rvert$ [$\times 10^3$]	Location
SEL-2018 (600 m)	SEL	✓	20	15 (5)	Selhausen, DE
WST-2019 (1500 m)	WST	✓	20	14 (5)	Braccagni, IT
CKA-2020 (600 m)	CKA	✓	20	10 (3)	Kl. Altendorf, DE
CKA-2020 (350 m)	CKA	✓	20	8 (2)	Kl. Altendorf, DE
CKA-2021 (350 m)	CKA	✓	20	4 (1)	Kl. Altendorf, DE
TOPO	SOP, HOE *600 m* *2021 - 2023*	–	300	11 (3)	Jülich, DE Hölstein, CH
PRE	PHY, HOE, TR32 CKA, SEL, SOP WST, NRS *350 - 1800 m* *2018 - 2023*	–	300	235 (38)	

Table 1: Data sets of compiled HyPlant acquisitions from different locations in the years 2018 - 2023. *Data Set* denotes a single compilation. With *Campaign* we denote the campaign identifier pointing to the used acquisitions according to the identified scheme outlined in the openly available HyData data set [30], with *FLOX* we denote the availability of simultaneous FLOX data, with Δh the maximum topographic variation over the compiled data set, with $\lvert\mathcal{D}\rvert$ the data set size in terms of number of 60×60 image crops. In parenthesis is reported the number of patches used for training.

SIF measurements, we are able to complement the present study with a direct comparison of SIF estimates of our approach with ground-based in-situ SIF estimates. However, we point out that while HyPlant is well suited to test the set-up presented in this work, EmSFMNN may be applied on data acquired by other airborne or spaceborne sensors. Its specific formulation is in fact well suited to cope with large existing hyperspectral data sets and continuous data streams of hyperspectral imaging sensors.

2. Data

2.1. Data quality provided by the HyPlant FLUO sensor

The HyPlant FLUO sensor [29] is the airborne demonstrator for the spaceborne FLEX satellite mission [22]. As such, it has been designed specifically for SIF retrieval in the atmospheric O_2-A and O_2-B absorption bands with a spectral sampling interval of 0.11 nm and a full width a half maximum (FWHM) of 0.25 nm. A large collection of hyperspectral HyPlant data sets

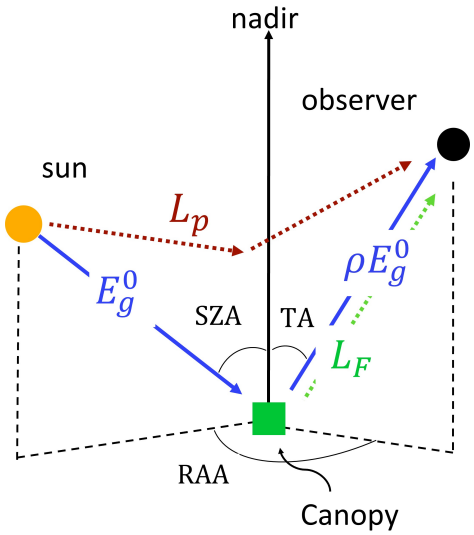


Figure 1: Geometrical set-up of the sun-observer geometry definitions in use. *RAA* denotes the relative azimuth angle, *TA* the tilt angle and *SZA* the sun zenith angle.

Specification		**Range**
Atmosphere	H_2O [cm]	0.3−3.0
	AOT_{550} []	0.02−0.30
Geometry	TA [°]	0−25
	SZA [°]	20−55
	RAA [°]	0−180
	h_{gnd} [km]	0−0.760
	h_{agl} [km]	0.2−2.86
Surface	ρ_{740} []	0.05−0.60
	s [nm^{-1}]	0−0.012
	e []	0−1
	F_{737} [*]	0−8
Sensor	$\Delta\lambda$ [nm]	$[-0.080, +0.080]$
	$\Delta\sigma$ [nm]	$[-0.040, +0.040]$
Input dimensions		13
Number of bands		349
Number of samples		6.3×10^6

Table 2: Specification of the ranges of all physical variables necessary for complete parametrization of the simulation tool.
*: F_{737} is given in units of [mW/nm/sr/m^2].

have been collected since 2014 [31, 32, 33, 34, 35, 36, 37] and are partly openly available [30]. In particular, yearly data sets since 2018 can be considered to be comparable across different campaigns due to their operationalized and standardized radiometric calibration and derivation of the geometric correction. Overall, the radiometric calibration is gauged at a mean relative uncertainty r_g of 3% [36] and the geolocalization reaches subpixel accuracy [29].

In this study, we make use of radiometrically corrected HyPlant FLUO acquisitions acquired in the years 2018 -2023 (cf. Tab. 1) in different flight campaigns, various locations and varying sun-observer geometries. The data set incorporates a large part of all available HyPlant FLUO acquisitions from this time period. We notably include acquisitions with strong topographic variation to train and to test the retrieval performance under these demanding conditions (cf. Sec. 4.4).

2.2. Simulation of HyPlant at-sensor radiance

The emulator utilized in this work is based on the polynomial emulator described in [27, 28] derived from a simulation tool generating single pixel at-sensor radiance [11]. It uses MODTRAN6 to model radiative transfer through the atmosphere and assumes simple parametric models for surface reflectance and fluorescence emission in the spectral range around the O_2-A oxygen absorption band. The parameters of the simulator have been chosen according to an extensive sensitivity study performed in [11]. We fixed the ranges of the resulting 13 parameters such that they cover the empirical distributions found in the HyPlant acquisitions used in this work (cf. Tab. 2). Appropriate ranges for the geometric parameters sensor altitude above ground level h_{agl}, ground altitude h_{gnd}, relative azimuth angle RAA and tilt angle TA could be established exactly from metadata provided with HyPlant data taking account of its precise orientation. The ranges for the surface parameters and for the sensor characterization also were adopted directly from preparatory work in [27, 28]. The parameterization of a simple quadratic reflectance model implemented in the simulation tool was chosen according to an analysis of vegetation and soil reflectance spectra of the DUAL hemispherical-directional reflectance product that is computed operationally for all HyPlant acquisitions. Equally, we modelled fluorescence emission in the O_2-A band spectral region with a Gaussian with fixed mean ($\mu = 737$ nm), fixed standard deviation ($\sigma_f = 20$ nm) and a free amplitude F_{737}. The ranges regarding the sensor characterization parameterized by center wavelength shifts $\Delta\lambda$ and FWHM shifts $\Delta\sigma$ were derived from in-flight data. Due to lacking simultaneous measurements, which would have allowed an estimate of the ranges of the atmospheric parameters aerosol optical thickness AOT_{550} and water vapour density H_2O, these ranges were chosen such that they covered all possible atmospheric states in which HyPlant campaigns are operated (cloud-free weather conditions in mid-latitude regions in summer).

We sampled the parameter ranges in Tab. 2 with different sampling strategies for training and validation data set to derive an emulator as outlined in [27, 28]. Importantly, the input parameters p were sampled independently. Since the parametric models for the spectral shapes of the reflectance and fluorescence implemented in the simulation tool were completely independent as well, we prevented our retrieval method to incorporate cross-correlations between fitted parameters a-priori as this would undermine the purely physical approach followed in this work.

2.3. In-situ SIF validation data

For a subset of the HyPlant acquisitions used in this work in-situ measurements of SIF are available (cf. Tab. 1). All in-situ measurements were derived with the Improved Fraunhofer Line Discrimination Method (iFLD) [38] from radiance point measurements and solar irradiance recordings of the hyperspectral FLOX device (Fluorescence Box, JB-Hyperspectral Devices GmbH, Duesseldorf, Germany). In the case of the FLOX measurement series matching with CKA-2020 HyPlant acquisitions, the in-situ measurements were taken in four different locations by four different devices. One FLOX was placed in an agricultural oat field and three others in wheat fields. For the validation, we have aggregated the time series and did not differentiate between the different FLOX devices. The localization of those FLOX systems was improved with an exact GPS RTK measurement at each of the devices. In case of the in-situ measurements matching the SEL-2018 HyPlant acquisitions, a single mobile FLOX device was used in agricultural fields of sugar beet and wheat.

FLOX measurements falling within a 5 minutes to the acquisition time of HyPlant were considered. We selected only FLOX measurements flagged as having high radiometric stability ($< 1\%$ difference in solar irradiance over the course of the measurement) in order to exclude measurements affected by cloud and haze. Since HyPlant campaigns are only conducted in optimal weather conditions, no measurements had to be excluded. In the case of multiple measurements within this time window matching a single acquisition, we averaged the FLOX iFLD SIF estimate to compare with HyPlant derived SIF estimates. In order to account for localization errors as well as the field of view we compared HyPlant pixels within a 2 m radius around the measurement location. The temporal and spatial variance resulting from the time windowing and spatial localization buffer were used as proxies for uncertainty estimates in the performance calculations.

3. Methods

3.1. Simulation Tool

The simulation tool utilized in this work [11, 27, 28] uses MODTRAN6 to simulate the HyPlant at-sensor radiance in a spectral range covering the

O_2-A absorption band (740 - 780 nm) according to the model

$$L_s(\mathbf{p}) = \left(L_p + \frac{E_g^0 \rho T^{\uparrow}}{\pi(1-\rho S)} + L_{\mathrm{F}} T^{\uparrow} \right)(\mathbf{p}) \tag{1}$$

as a function of the parameters $\mathbf{p}$ (as defined in Tab. 2) where L_p is the path radiance, E_g^0 is the global solar irradiance on the ground, $T^{\uparrow}$ is the total transmission coefficient from surface to sensor (direct and diffuse components, $T^{\uparrow} = T^{\uparrow}_{\mathrm{dir}} + T^{\uparrow}_{\mathrm{dif}}$), L_F is the top-of-canopy fluorescence emission modelled as a Gaussian with fixed variance and amplitude F_{737}, S is the spherical albedo of the atmosphere, ρ is the hemispherical-directional reflectance modelled as a 2^{nd} order polynomial with offset ρ_{740}, slope s and curvature e, H_2O denotes the columnar water vapour content, AOT_{550} the aerosol optical thickness, TA the viewing angle, SZA the solar zenith angle, RAA the relative azimuth angle between observation and irradiance directions, h_{gnd} the topographic height above sea level, h_{agl} the sensor height above ground level and $\Delta\lambda$ and $\Delta\sigma$ the scalar shifts in the center wavelength and full width at half maximum (FWHM). In order to be consistent with in-situ reference fluorescence estimates, we report SIF_{760} instead of F_{737} in all validation and analysis sections which we define as the functional value of the modelled fluorescence emission at 760 nm. The simulations have been conducted with an atmospheric model corresponding to the MODTRAN *mid-latitude summer* model. Thus, we have disregarded changes in the atmospheric pressure profile that might be caused by changing meteorology or topography. Finally, as in [28, 27], we densely sample the parameter space spanned by the parameter ranges in Tab. 2 and run a total of 6.3×10^6 simulations.

3.2. Definition of the polynomial emulator

Emulation of a hyperspectral simulator $L_s(\mathbf{p}) : \mathbb{R}^M \rightarrow \mathbb{R}^\Lambda$ from physical parameters $\mathbf{p} \in \mathbb{R}^M$ by an emulator e is ultimately a regression problem where we derive a function $e : \mathbb{R}^M \rightarrow \mathbb{R}^\Lambda$ that reproduces as closely as possible the simulator L_s at reduced computational cost. In practice, there is a trade-off between reducing the residual between simulator and emulator on the one hand and reducing the computational cost of e on the other for any non-trivial simulator L_s. Since the emulator is used during the training of a neural network, we require additionally that its gradient computation is efficient and preferably can be integrated easily in common programming frameworks for

deep learning. The polynomial emulator investigated by [27, 28] fulfills these requirements. It is defined as the polynomial function of dth order

$$e_d(\mathbf{p} \,|\, \mathbf{a}_{\mathcal{K}^d}) = \sum_{\mathbf{k} \in \mathcal{K}^d} p_1^{k_0} \dots p_M^{k_M} \mathbf{a}_\mathbf{k}, \tag{2}$$

over the parameters $\mathbf{p}$, where $\mathbf{a}_\mathbf{k} \in \mathbb{R}^\Lambda$ and where the set of polynomial features is defined as

$$\mathcal{K}^d = \left\{ \mathbf{k} \in \mathbb{N}^M : \sum_{1 \le i \le M} k_i \le d \right\} \tag{3}$$

We train the emulator weights $\mathbf{a}_\mathbf{k}$ in a least-squares optimization with a training subset of the data points partitioned from the total simulation data set as in [27, 28].

3.3. Emulation of a wavelength dependent sensor characterization

The simulation data base is created for scalar shifts $\Delta\lambda$ and $\Delta\sigma$, i.e. simulated spectra $s(\mathbf{p}) \in \mathbb{R}^\Lambda$ will suffer the same simulated sensor miscalibration in all wavelengths $\lambda \in \Lambda$. In a realistic sensor model, CW and FWHM shifts are, however, functions of the wavelength such that we ought to find an emulator with dependency on shifts $\mathbf{\Delta}\lambda \in \mathbb{R}^\Lambda$ and $\mathbf{\Delta}\sigma \in \mathbb{R}^\Lambda$ in addition to the other input parameters $\tilde{\mathbf{p}}$. We assume that there is no cross dependency of the shifts either in the measured at-sensor radiance L or the simulator L_s, i.e.

$$\forall i \neq k : \; \frac{dL_i}{dv_k} = \frac{d\left(L_s\right)_i(\tilde{\mathbf{p}}, \mathbf{\Delta}\lambda, \mathbf{\Delta}\sigma)}{dv_k} = 0, \; v \in \{\mathbf{\Delta}\lambda, \, \mathbf{\Delta}\sigma\}. \tag{4}$$

In this case a naive approach to extend the emulator could be achieved by rewriting

$$e_d^\Lambda(\tilde{\mathbf{p}}, \mathbf{\Delta}\lambda, \mathbf{\Delta}\sigma) = \left(e_d(\lambda_i \,|\, \tilde{\mathbf{p}}, \Delta\sigma_i, \Delta\lambda_i)\right)_{0 \le i \le \Lambda} \tag{5}$$

As the simulation data base covers a large number of spectral bands ($\Lambda = 349$) such an approach results in a significant increase in computation time for a single spectrum since the emulator would need to be run Λ times for a single emulated spectrum. We therefore adopt an approximation. We derive a multiplicative correction factor

$$m(\lambda_i \,|\, \mathbf{\Delta}\lambda, \mathbf{\Delta}\sigma) = \mathbb{E}\left[w\left(\lambda_i \,|\, \tilde{\mathbf{p}}, \, \Delta\lambda_i, \Delta\sigma_i\right)\right] \tag{6}$$

$$= \mathbb{E}\left[\frac{e_d\left(\lambda_i \,|\, \tilde{\mathbf{p}}, \, \Delta\lambda_i, \Delta\sigma_i\right)}{e_d\left(\lambda_i \,|\, \tilde{\mathbf{p}}, \, \Delta\lambda_i = \Delta\sigma_i = 0\right)}\right] \tag{7}$$

Module	Parameters	
Encoder e_{in}	*Dim.*	(2e3, 2e3, 1e3, 5e2, 5e2, 1e2, 1e2, 1e2, 50)
	Reps.	(3, 3, 3, 3, 3, 3, 3, 1, 1)
	D_p	(0.05, 0.05, 0.01, 0.01, 0.005, 0.0)
Decoder d_v	*Dim.*	(1e2, 50, 50, 50, 10)
	Reps.	(3, 2, 2, 1)
	D_p	(0.05, 0.05, 0.01, 0.01, 0.005, 0.0)
Sensor charact. g	*Dim.*	(1e2, 50, 50, 50, 10)
	Reps.	(3, 2, 2, 1)
	D_p	(0.05, 0.05, 0.01, 0.01, 0.005, 0.0)

Table 3: Dimensionalities for different modules in the EmSFMNN architecture (cf. Fig. 2). Elements in a tuple denote an architecture parameter for a single sublayer in a module. *Reps.* denotes the number of repetitions of linear layers in a sublayer, D_p denotes the dropout rate of the output of the sublayer. For a more detailed exposition of the module architecture we refer to [24].

where the expectation is calculated by sampling randomly over the parameter distribution in the input parameter space. As we will show below, the variance over this distribution is very small, such that we can write

$$\tilde{e}_d\left(\lambda_i \,|\, \tilde{\mathbf{p}}, \boldsymbol{\Delta}\lambda, \boldsymbol{\Delta}\sigma\right) \approx m(\lambda_i \,|\, \boldsymbol{\Delta}\lambda, \boldsymbol{\Delta}\sigma) \cdot e_d\left(\lambda_i \,|\, \tilde{\mathbf{p}},\, \Delta\lambda_i = \Delta\sigma_i = 0\right) \quad (8)$$

3.4. Neural Network Architecture

We integrate a fourth-order emulator $\tilde{e}_4$ of the radiative transfer in Eq. 1 with a self-supervised neural network. We construct a neural network acting on fixed size excerpts of HyPlant imagery (60×60) that we will refer to as *patches*. The network architecture is defined as in SFMNN [24] and similarly to a DESIS EmSFMNN implementation [25] (cf. Fig. 2) as a Multilayer Perceptron (MLP) encoder-decoder set-up. The network is trained to predict all parameters $\mathbf{p}$ of the RTM model in Eq. 1 that cannot be inferred from metadata or geometrical recordings, i.e. all parameters in Tab. 2 except parameters of the group *Geometry*.

The encoder e_{in} and decoder modules d_v in this network are constructed as MLPs with residual links and have the dimensionalities given in Tab. 3. The decoders are tasked with disentangling the latent space spanned by the encoder to the physical parameters $\tilde{\mathbf{p}}$ parameterizing the radiative transfer

model underlying the simulation tool and, thus, the emulator. We define two decoders d_v: for the reflectance and fluorescence related parameters predicted for each pixel and for the atmospheric parameters predicted for each patch. These two decoder modules are implemented identically with the exception of a final spatial mean reduction before the emulator layer in the case of the patchwise predictor. As in SFMNN, we differentiate between pixel-wise and patch-wise prediction based on the fact that atmospheric parameters (AOT and H_2O) have an autocorrelation that typically exceeds the physical patch size such that a single atmospheric estimate per patch can be assumed to lead to sufficiently precise approximations.

The estimation of sensor shifts $\mathbf{\Delta}\lambda$ and $\mathbf{\Delta}\sigma$ is implemented differently. We assume that we can fit these shifts as a function of the sensor state at acquisition time and the across-track sensor position alone without any spectral input. This assumption is implemented in the architecture by estimating the sensor shifts only from an arbitrarily defined acquisition identifier $\mathbf{u} \in \mathbb{R}^U$ that represents the sensor state and the across-track position x_1. At the start of the training we randomly instantiate these identifiers $\mathbf{u}$ of fixed dimensionality ($U = 8$) for each acquisition in the training data set and include them as learnable parameters in the optimization. The MLP module g predicts shifts $\mathbf{\Delta}\lambda$ and $\mathbf{\Delta}\sigma$ for each wavelength ($\Lambda = 349$) at across-track positions x_1 from pixelwise concatenations of the identifier vectors $\mathbf{u}$ and a positional encoding of $\mathbf{x}_1$ [39].

An important characteristic of this particular set-up consists in the physically coherent separation of inputs and the differentiation of output dimensions for individual parameters. For example, all reflectance parameters (ρ_{740}, s, e) and the fluorescence emission amplitude F_{737} are estimated for each pixel from the radiance data and geometrical information ν_{geo}, but without providing the acquisition identifier u since the decoders to those parameters by definition do not depend on sensor characteristics or acquisition dependent changes. Similarly, atmospheric parameters are estimated from radiance and ν_{geo} alone, but, differently to the surface parameters, only per patch as we assume negligible variance of these parameters over small spatial distances. The sensor characterization $\Delta\lambda$ and $\Delta\sigma$ on the other hand is uniquely estimated from the acquisition identifier u for individual across-track positions x_1 since it is driven by factors that are identical across single acquisitions. Both input separation and differentiation in output dimensionality constrain the network optimization architecturally with prior knowledge of the physical processes and sensor design at play. On the other hand, we implicitly con-

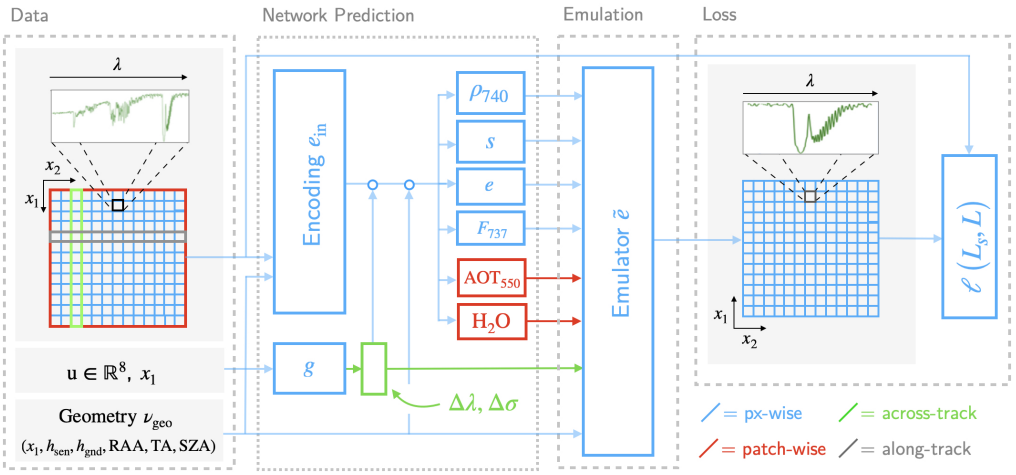


Figure 2: Outline of the architecture and emulator integration of the SFMNN used in this work. Dimensions of the encoder e_{in}, the decoders for variables ρ_{740}, s, e, F_{737}, AOT_{550} and H_2O and the sensor characterization g are given in Tab. 3.

strain the network by enforcing physically accurate solutions of the radiative transfer equation Eq. 1 given a particular parametrization $\tilde{\mathbf{p}}$. Differently to the simplified four-stream model used in SFMNN to model at-sensor radiances, the emulator $\tilde{e}$ allows for pixel-wise parameterization of the radiative transfer formulation with known geometrical variables. This is a significant improvement over SFMNN's formulation as the solution space of the network can be constrained very precisely in a pixel-wise fashion.

3.5. Loss formulation

The loss used in this set-up is adapted from the loss used in [24]. It consists of a batchwise mean squared reconstruction error complemented by two regularizers. Given the input radiance spectra L_{HyP} as measured by HyPlant and matching geometrical meta data ν_{geo} (flight h_{sen} and ground altitude h_{gnd}, relative azimuth RAA, tilt angle TA and solar zenith angle SZA) we train the network n to minimize

$$\ell\left(L_{\text{HyP}}, \hat{L}_{\text{HyP}}\right) = \left\langle \left(L_{\text{HyP}} - \hat{L}_{\text{HyP}}\right)^2 \right\rangle_{\lambda,\, x} + \gamma_f\, \ell_f + \gamma_N\, \ell_{\text{NDVI}}, \tag{9}$$

where L_{HyP} is the measured at-sensor radiance in the spectral window $\mathcal{W}$ and

$$\hat{L}_{\text{HyP}} = \tilde{e}(\tilde{\mathbf{p}}, \boldsymbol{\Delta}\lambda, \boldsymbol{\Delta}\sigma, \nu_{\text{geo}}) \tag{10}$$

denotes the network prediction with predicted $\tilde{\mathbf{p}}$, $\mathbf{\Delta}\lambda$ and $\mathbf{\Delta}\sigma$. $\langle \dots \rangle_{x,\lambda}$ denotes the spatial and spectral mean over the patches included in a batch. Throughout all experiments in this work we have fixed the regularizer weights $\gamma_f = 1$ and $\gamma_{\mathrm{NDVI}} = 10$. These weights were established as they have shown satisfactory results in preliminary tests on the CKA-2020 (600 m) data set (cf. Tab. 1).

The fluorescence regularization

$$\ell_f = \left\langle \sum_{\lambda \in \mathcal{W}} w_\lambda \left(L_{\mathrm{HyP}}(\lambda) - \hat{L}_{\mathrm{HyP}}(\lambda) \right)^2 \right\rangle_x \Bigg|_{\delta p_i = 0,\, p_i \neq F_{737}} \tag{11}$$

boosts the contribution of reconstruction residuals according to a SNR-based weighting w_λ that accounts for the spectral distribution of typical fluorescence emission. This weighting is derived as the Moore-Penrose solution to a linearized retrieval problem with known reflectance and atmospheric parameters [24]. We thus restrict the gradient contribution of this loss term to affect only the fluorescence decoder (i.e. network weights uniquely related to parameters $p_i \neq F_{737}$, i.e. reflectance, atmospheric and sensor parameters, are not affected by this term). The physiologically motivated regularizer

$$\ell_{\mathrm{NDVI}} = \left\langle \hat{f} \cdot \delta \left(\mathrm{NDVI} < \tau \right) \right\rangle_x \tag{12}$$

ensures that the fluorescence estimate f vanishes in pixels with very low green vegetation, i.e. in pixels with a low Normalized Difference Vegetation Index (NDVI). To identify these pixels, we set a threshold $\tau = 0.15$ on an approximate NDVI product derived from the radiance L_{HyP}.

3.6. Training set-up

The training of the EmSFMNN SIF predictors takes place in two steps. We first train a backbone on the PRE HyPlant data set (cf. Tab. 1). This backbone is used as the initialization to all EmSFMNN instances that are trained for individual data sets in the second step. Finetuning of PRE aims at adjusting the network (1) to the data set specific radiance calibration, (2) to train the acquisition specific identifiers u that determine the estimated shifts $\Delta\lambda$ and $\Delta\sigma$ and (3) to train in the specific parameter ranges covered differently in the various data sets (e.g. TOPO exhibiting larger variation of h_{agl}). During the finetuning step, the encoder e_{in} is fixed and only the decoders d_v, the identifiers u and the sensor characterization g are trained.

As the identifiers u are not estimated from the radiance data but rather implicitly as a result of the architectural constraint in each acquisition, there is no generalization property of this part of the network. Thus, in order to get meaningful shift predictions, a finetuning is thus necessary. When EmSFMNN instances are applied to data sets for which they were not finetuned, arbitrary identifiers u are selected from the set of finetuned u. This procedure results in larger reconstruction errors than would have been possible with a finetuning of u but it doesn't necessarily affect the fluorescence estimate negatively. The spectral reconstruction window $\mathcal{W}$ was fixed to cover 750 - 770 nm.

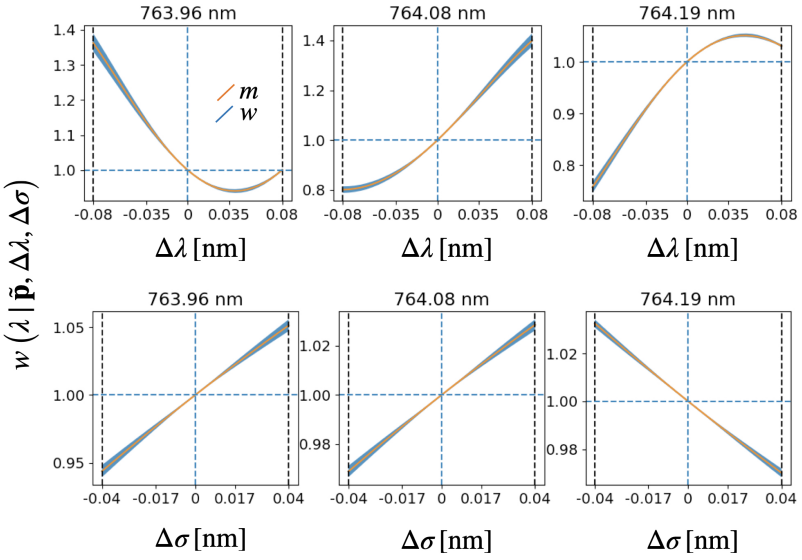


Figure 3: Multiplicative change w of e_4 under variable sensor shifts ($\Delta\lambda$ and $\Delta\sigma$) in three selected wavelengths. In blue is plotted the standard deviation of w (as defined in Eq. 6) over the distribution of randomly sampled emulator parameter configurations $\mathbf{p}$. The fitted mean used as multiplicative correction m (see Eq. 6) is plotted in orange.

4. Results

4.1. Training of emulator extension for bandwise spectral shifts

We have derived a polynomial emulator of 4^{th} order of HyPlant at-sensor radiance e_4 for the parameter ranges given in Tab. 2. In order to allow for

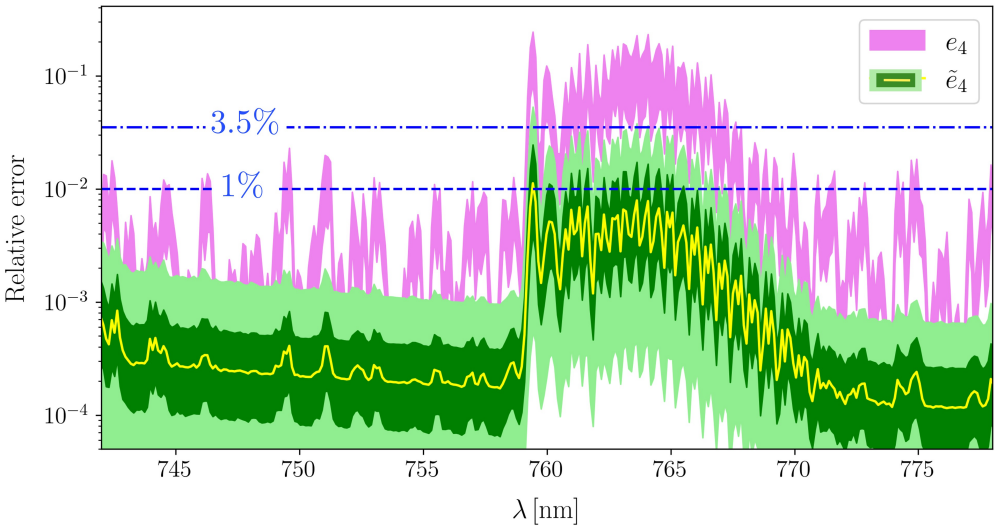


Figure 4: Relative errors of the approximate emulator $\tilde{e}_4$ (green/yellow) and scalar shift emulator e_4 (pink) with respect to the accurate e_4^Λ emulator. In the case of e_4 we set $\Delta\lambda = \Delta\sigma = 0$. The 25 -75 % percentile range is plotted in dark green, the 5 - 95 % percentile range in light green, the mean in yellow. Pink denotes the 25 -75 % percentile range.

efficient training we then have implemented the emulator extension for wavelength dependent shifts $\tilde{e}_4$ based on the polynomial emulator e_4 which acts only on scalar shifts as outlined in Sec. 3.2. To this end, we have computed the multiplicative factor m as the expectation in Eq. 6. To compute the distribution, we uniformly sampled a large number of parameter combinations $\mathbf{p}$ and sensor shifts $\boldsymbol{\Delta}\lambda$ and $\boldsymbol{\Delta}\sigma$ in the input space spanned by the individual parameter ranges.

We found the standard deviations of w to be bounded by 3.5% under CW shifts and 0.06% by FHWM shifts which we regarded as sufficiently small to approximate it by its mean m (cf. Fig. 3). Subsequently, we fitted a 5^{th} order polynomial to the derived m to gain a multiplicative factor defined on the whole input parameter space discarding the need for interpolation during prediction. The dimension of this polynomial was required to be just large enough to fit m well. The use of $\tilde{e}_4$ leads to a significant time reduction as compared to e_4^Λ (cf. Tab. 4).

In order to evaluate the accuracy of $\tilde{e}_4$ we compared it to e_4^Λ on a uniformly sampled test set. While e_4^Λ takes significantly longer to compute, its accuracy

with respect to the RTM is as high as the emulator itself since it essentially computes the emulator in a bandwise fashion. In Fig. 4 we show that the mean relative error incurred by using the approximation $\tilde{e}_4$ is smaller than 1% . However, the 95% percentile reaches a relative error of 3.5% inside the O_2-A band. We equally show the effect of neglecting bandwise shifts by comparing emulations of e_4^Λ with e_4 emulations with scalar shifts. The same parameters $\tilde{\mathbf{p}}$ were used for e_4^Λ and e_4 with only $\Delta\lambda$ and $\Delta\sigma$ set to a fixed scalar value for e_4. The relative errors can reach up to 10-20 % in the O_2-A band highlighting the importance of bandwise sensor characterization.

4.2. Reconstruction Performance

We evaluate the impact of the various implemented constraints, the optimization and the emulator extension on the reconstruction performance. To this end we compare the reconstruction performance of four different EmSFMNN set-ups to the reconstruction performance of an unconstrained least-squares optimization (LSQ) of the emulator e_4 to individual pixels in a single HyPlant acquisition (cf. Fig. 5). By $m_{\mathrm{CKA}}(\tilde{\mathrm{e}}_4)$ we denote an EmSFMNN predictor using the spectrally explicit sensor miscalibration emulator $\tilde{e}_4$ and finetuned on the CKA-2020 (600 m) data set. The EmSFMNN $m_{\mathrm{SEL}}(\tilde{\mathrm{e}}_4)$ and $m_{\mathrm{SEL}}(\mathrm{e}_4)$ are equivalently trained on the SEL-2018 (600 m) data set and $m_{\mathrm{PRE}}(\tilde{\mathrm{e}}_4)$ denotes the common backbone without finetuning. Importantly, the HyPlant acquisition for which we evaluate the reconstruction performance is part of the finetuning training set of $m_{\mathrm{SEL}}(\tilde{\mathrm{e}}_4)$ but not of $m_{\mathrm{CKA}}(\tilde{\mathrm{e}}_4)$. Due to the prediction of $\Delta\lambda$ and $\Delta\sigma$ in EmSFMNN being dependent on learnable IDs, and the ID not having been trained for $m_{\mathrm{CKA}}(\tilde{\mathrm{e}}_4)$, we use a single ID in CKA that we arbitrarily choose from the set of IDs trained for CKA acquisitions.

	e_4	e_4^Λ	$\tilde{e}_4$
Prediction time per sample	0.28 μs	55.40 μs	1.93 μs

Table 4: Prediction time measurements for the original emulator e_4, the original emulator applied in a bandwise fashion e_4^Λ and the emulator approximation $\tilde{e}_4$. In the case of e_4 only scalar sensor shifts were computed. The values represent the average of 20 time measurements on a single GPU (NVIDIA Quadro RTX 8000) predicting a batch of 10^4 samples.

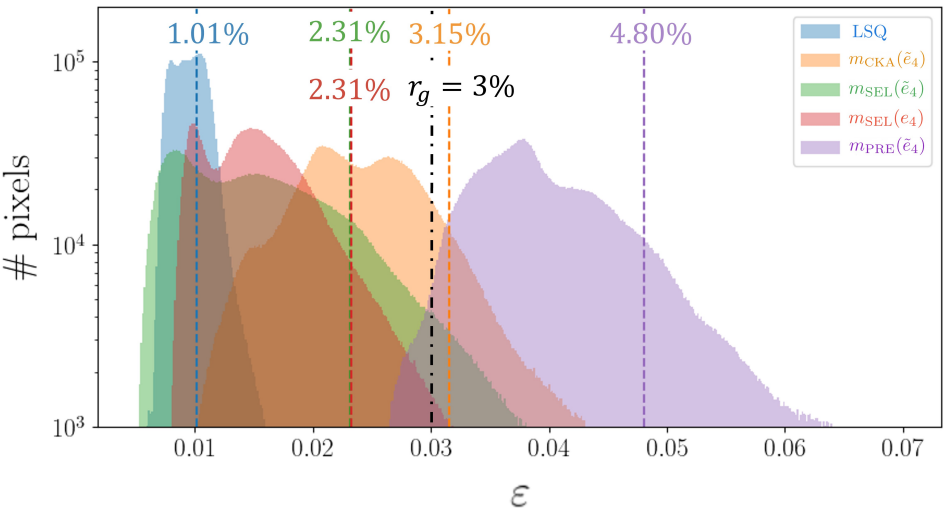


Figure 5: Distribution of relative reconstruction errors ε (see Eq. 13) for different EmSFMNN set-ups as well as a least-squares optimization in a HyPlant acquisition recorded on 2018/07/26 15:30 CEST in Selhausen. The mean values of these distributions are reported directly in the figure with vertical lines. Mean relative reconstruction errors for $m_{\mathrm{SEL}}(\mathrm{e}_4)$ and $m_{\mathrm{SEL}}(\tilde{\mathrm{e}}_4)$ overlap in this figure. r_g denotes the relative uncertainty of the radiometric calibration.

In Fig. 5 we report the distribution of the relative reconstruction residuals

$$\varepsilon = \left\langle \left| \frac{L_{\mathrm{HyP}} - \hat{L}_{\mathrm{HyP}}}{L_{\mathrm{HyP}}} \right| \right\rangle_{x,\lambda} . \tag{13}$$

of LSQ and all EmSFMNN predictors in the acquisition. We find that the unconstrained emulator optimization LSQ outperforms all EmSFMNN training set-ups. However, since LSQ is completely unconstrained, the resulting estimates of physical parameters defining the simulation layer are not well disentangled. As a consequence, it cannot be used for SIF retrieval, even though it provides a useful baseline for the reconstruction error.

Since the simulation layer of $m_{\mathrm{SEL}}(\mathrm{e}_4)$ and LSQ are the same, a comparison of $m_{\mathrm{SEL}}(\mathrm{e}_4)$ to the least-squares optimization LSQ isolates the impact of EmSFMNN's constraint formulation and its feature-based optimization. The direct EmSFMNN equivalent $m_{\mathrm{SEL}}(\mathrm{e}_4)$ performs significantly worse than LSQ, presumably due to the constrained optimization. However, this decrease in reconstruction performance can be improved by adopting the ex-

		r	$\mathbf{R}^2$	MAE [mW nm^{-1} sr^{-1} m^{-2}]	N
SEL-2018 (600 m)	EmSFMNN ($\tilde{e}_4$)	0.91	0.55	**0.26 ± 0.09**	10
	EmSFMNN (e_4)	0.86	**0.74**	0.46 ± 0.05	10
	PRE	0.78	0.57	0.47 ± 0.06	10
	SFMNN	**0.98**	–	0.68 ± 0.07	10
	SFM	0.96	–	0.51 ± 0.07	10
	iFLD	0.64	0.10	0.88 ± 0.00	11
WST-2019 (1500 m)	EmSFMNN ($\tilde{e}_4$)	**-0.54**	–	**0.29 ± 0.05**	15
	EmSFMNN (e_4)	-0.73	–	1.48 ± 0.05	15
	PRE	-0.78	–	0.41 ± 0.04	15
	SFMNN	–	–	**0.22 ± 0.10**	15
	SFM	–	–	0.53 ± 0.08	15
	iFLD	–	–	0.80 ± 0.10	15
CKA-2020 (600 m)	EmSFMNN ($\tilde{e}_4$)	0.65	0.02	**0.35 ± 0.05**	16
	EmSFMNN (e_4)	0.69	0.18	0.47 ± 0.05	16
	PRE	0.67	–	0.39 ± 0.06	16
	SFMNN	0.69	**0.34**	**0.33 ± 0.06**	16
	SFM	**0.72**	–	0.48 ± 0.06	16
	iFLD	0.64	–	0.42 ± 0.09	16
CKA-2020 (350 m)	EmSFMNN ($\tilde{e}_4$)	0.74	0.04	**0.28 ± 0.04**	34
	EmSFMNN (e_4)	0.81	**0.12**	0.35 ± 0.04	34
	PRE	0.80	–	0.33 ± 0.04	34
	SFMNN	0.84	–	0.34 ± 0.04	34
	SFM	**0.87**	–	0.35 ± 0.04	34
	iFLD	0.58	0.05	**0.28 ± 0.05**	34
CKA-2021 (350 m)	EmSFMNN ($\tilde{e}_4$)	–	0.19	0.38 ± 0.09	6
	EmSFMNN (e_4)	–	0.16	1.07 ± 0.12	6
	PRE	–	0.30	0.70 ± 0.09	6
	SFMNN	–	–	0.65 ± 0.10	6
	SFM	–	–	0.50 ± 0.08	6
	iFLD	**0.85**	**0.71**	**0.12 ± 0.18**	6

Table 5: Comparative validation of SFM, iFLD, SFMNN and EmSFMNN retrieval methods. We report the mean absolute error (MAE) of the EmSFMNN predictions with respect to FLOX measurements, the Pearson correlation r and the Explained Variance Score R^2. In cases where the p-value of r is larger than 5% we do not report r and write $-$ instead. Similarly, in cases where $R^2 \leq 0$ we do not report R^2 and write $-$.

tended emulator formulation with bandwise shift prediction which results in a smaller lower limit of the reconstruction residuals but which also increases the tail towards larger residuals.

We measure the generalization capacity of EmSFMNN across different data sets with respect to the reconstruction performance. The residual distribution of $m_{\mathrm{CKA}}(\mathrm{e}_4)$ yields an ε which is significantly increased over its equivalent $m_{\mathrm{SEL}}(\mathrm{e}_4)$ as it lacks finetuning to the HyPlant input data acquisition. However, its performance is close to what can be expected from the relative uncertainty of the radiometric calibration of the at-sensor radiance r_g. Since $m_{\mathrm{CKA}}(\mathrm{e}_4)$ that has been finetuned on the CKA-2020 data set is applied here on an acquisition from the SEL-2018 data set, effects due to different yearly calibrations can affect the reconstruction performance.

Finally, it can be observed that the relative reconstruction error of $m_{\mathrm{PRE}}(\mathrm{e}_4)$ is constrained in the range 3 - 6 %. It is thus larger than errors attributed purely to calibration uncertainties and indicates that the backbone PRE is not able to reconstruct model at-sensor radiance of arbitrary HyPlant acquisitions from its learned feature representation without prior finetuning.

4.3. Validation with FLOX data

We validate EmSFMNN SIF predictions with top-of-canopy iFLD SIF estimates derived from radiance data recorded by FLOX devices. To this end, we use the fluorescence model assumption of the Gaussian implemented in the simulations to calculate SIF_{760} consistent with the FLOX iFLD retrieval software. Five measurement time series are at our disposal acquired during HyPlant overflights in field campaigns in the years 2018 - 2021 (see Fig. 6). To support our comparison, we also report the validation results for three baseline methods that were developed for hyperspectral HyPlant imagery (iFLD, SFM and SFMNN). Additionally, we show the impact of the emulator formulation and the finetuning on the performance of the standard training set-up denoted by EmSFMNN ($\tilde{e}_4$) in Tab. 5. To this end, we report (i) results for the EmSFMNN set-up using the polynomial interpolation without band-wise sensor characterization, denoted as EmSFMNN (e_4), and (ii) the performance of the coarsely pre-trained SIF predictor, denoted as PRE.

We find that the EmSFMNN ($\tilde{e}_4$) predictors finetuned to the individual datasets generally are among the best SIF retrieval methods in terms of the mean absolute error with respect to FLOX estimates (MAE). They yield MAE scores consistently smaller than 0.4 mW nm^{-1} sr^{-1} m^{-2} whereas stronger variation in MAE can be found in case of the iFLD, SFM and

SFMNN SIF predictions. We find notably a reduced overestimating bias of EmSFMNN as compared to SFMNN in Fig. 6 and a higher accuracy than in SFM and iFLD predictions.

Both the use of $\tilde{e}_4$ instead of e_4 as well as the finetuning considerably decrease the MAE. This can be concluded from the improved performance of EmSFMNN ($\tilde{e}_4$) over EmSFMNN (e_4) and PRE. These two aspects have been introduced to improve
the model representation by (i) a more exact simulation layer in the reconstruction loss and (ii) a more precise fitting of the spectral data by specializing the network weights to a selection of HyPlant lines. While the use of the non-specialized pretrained EmSFMNN model PRE leads to validation results comparable to the baseline methods, it is necessary to make use of a finetuning step to improve over the baseline methods iFLD, SFM and SFMNN.

In Tab. 5 we equally report the Pearson correlation scores r and the Explained Variance Score R^2. A consistent cross-validation assessment based on r and R^2 is, however, not possible throughout all validation data sets due to strongly variable performance under these two metrics. Due to the small number of validation data points these performance metrics are subject to large uncertainties. A reduced performance of EmSFMNN ($\tilde{e}_4$) in terms of r with respect to SFMNN is, however, observable in all data sets where such a comparison can be made. We hypothesize that this may be due to the more strict modelling approach of EmSFMNN which may result in a higher sensitivity to sensor noise.

4.4. Topography

The emulator was derived from simulations covering a h_{gnd} range of 0 - 0.76 km and of h_{agl} 0.2 - 2.86 km. This allows the application of EmSFMNN predictors in acquisitions with large height variation where both the surface height h_{gnd} and flight height above ground level h_{agl} change significantly over the course of a single datatake. To test the reconstruction performance of EmSFMNN predictors under these circumstances we examine the TOPO data set consisting of HyPlant acquisitions with strong topographic variation and a nominal flight height of 600 m (see Tab. 1). We apply (i) the EmSFMNN predictor finetuned to the CKA-2020 (600 m) data set (denoted as m_{CKA}) and (ii) a EmSFMNN finetuned to the TOPO data set (m_{TOPO}). The finetuning of the m_{TOPO} and m_{CKA} was performed on the TOPO data set as described above (cf. Tab. 1) and only differed in the finetuning data

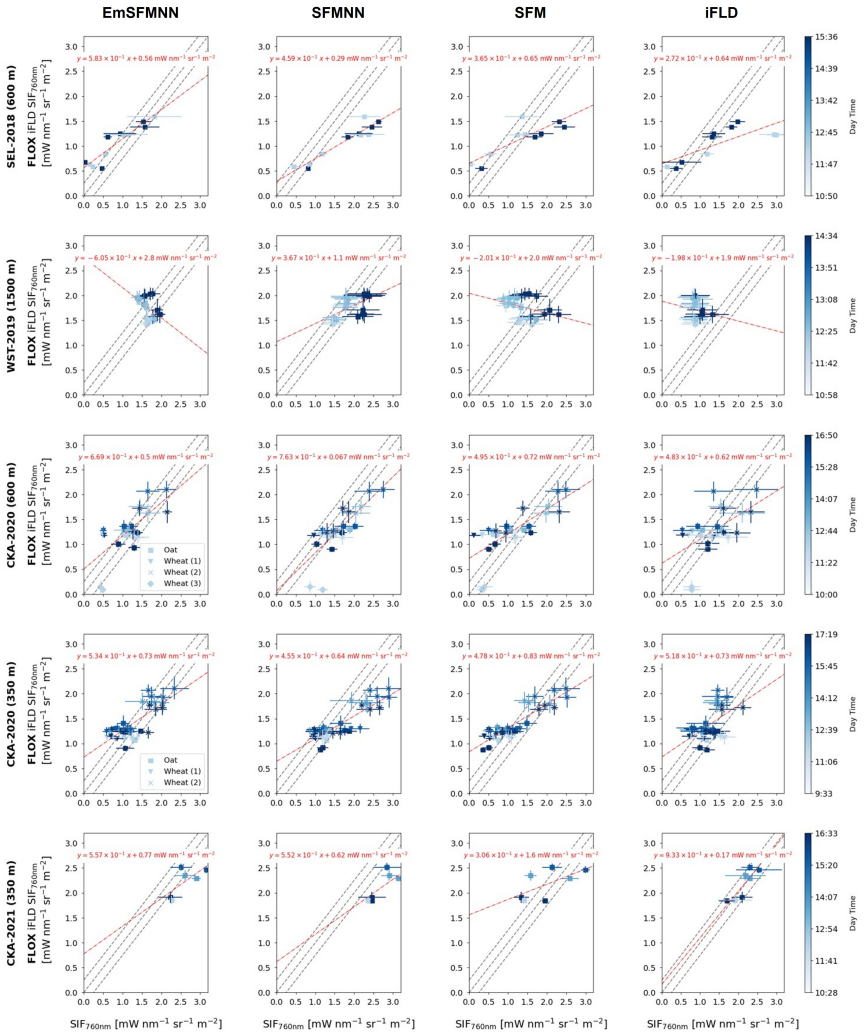


Figure 6: FLOX derived iFLD SIF vs. HyPlant derived EmSFMNN, SFMNN, SFM and iFLD SIF in the five in-situ validation data sets (see Tab. 1). The dashed line and the red floating labels report the linear relationship between EmSFMNN and FLOX iFLD estimates. In the CKA-2020 data sets FLOX measurements from different devices are reported separately.

set. Both predictors derive from a EmSFMNN backbone trained on the PRE data set which includes the data contained in TOPO. By examining

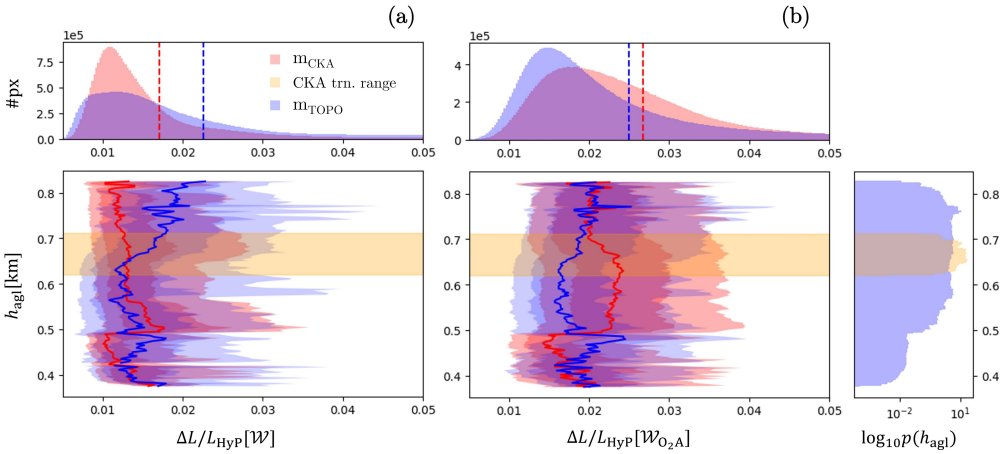


Figure 7: Top row: Marginalized distribution of relative reconstruction residuals of CKA(1) in red and TOPO(1) in blue in the whole fitted spectral window $\mathcal{W}$ (a) and in the spectral window $\mathcal{W}_{O_2A}$ in the O_2-A band (b). Bottom left: Relative reconstruction residuals of CKA(1) (red) and TOPO(1) (blue) stratified by the sensor height above ground h_{agl} in $\mathcal{W}$ (a) and $\mathcal{W}_{O_2A}$ (b). Red and blue lines denote the means, dark areas denote the 25 - 75 percentile ranges, light areas denote the 10 - 90 percentile ranges. In yellow we highlight the range of h_{agl} covered by the CKA-2020 data set on which CKA(1) was finetuned. Bottom right: Logarithmic empirical histogram of h_{agl} of CKA-2020 in yellow and TOPO in blue.

on TOPO with both m_{CKA} and m_{TOPO} we can evaluate the importance of topography related distribution differences between finetuning data sets.

In Fig. 7 (a) and (b) we summarize the residual statistics of m_{CKA} and m_{TOPO} as a function of the flight height h_{agl} in the full prediction spectral window $\mathcal{W}$ (750 - 770 nm) as well as in a narrow spectral window $\mathcal{W}_{O_2A}$ in the O_2-A absorption band (759.5 - 761 nm). m_{CKA} outperforms the finetuned m_{TOPO} in $\mathcal{W}$ exhibiting a residual distribution with less outliers. Notably, m_{CKA} outperforms m_{TOPO} including in h_{agl} ranges that are not covered by the CKA-2020 (600 m) finetuning data set. The finetuning to the validation data set TOPO yields, however, to an improved m_{TOPO} performance in the O_2-A band with m_{TOPO} exhibiting an improved reconstruction performance overall. The strong reconstruction residual outliers of m_{TOPO} are consequently contained in spectral regions outside the O_2-A band as can be understood from the fact that its performance on $\mathcal{W}_{O_2A}$ is less affected by it.

While we are able to assess the reconstruction performance of m_{TOPO} and m_{CKA}, we can not evaluate the SIF predictions in the TOPO data set

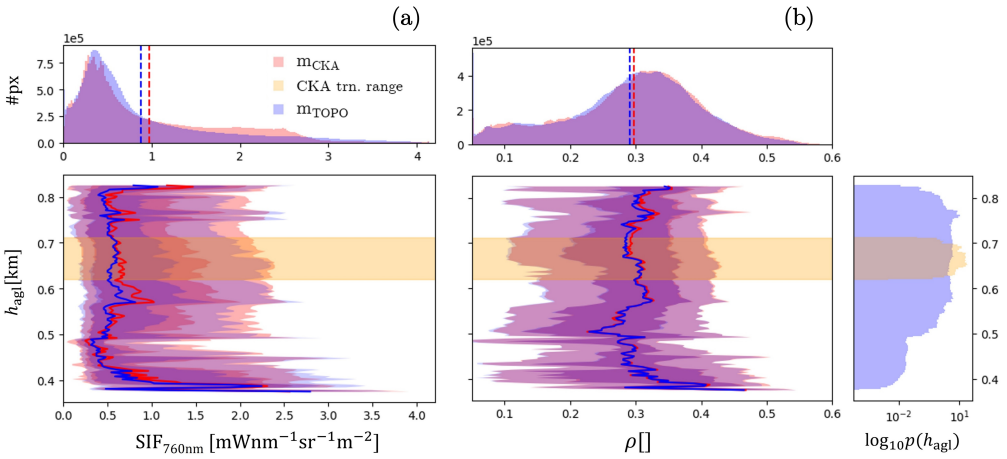


Figure 8: Top row: Marginalized distribution of reflectance offset parameter ρ (a) and fluorescence emission amplitude F_{737} as predicted by CKA(1) in red and TOPO(1) in blue. Bottom left: reflectance offset parameter ρ (a) and fluorescence emission amplitude F_{737} (b) stratified by the sensor height above ground h_{agl} as predicted by CKA(1) in red and by TOPO(1) in blue. Red and blue lines denote the means, dark areas denote the 25 - 75 percentile ranges, light areas denote the 10 - 90 percentile ranges. In yellow we highlight the range of h_{agl} covered by the CKA-2020 data set on which CKA(1) was finetuned. Bottom right: Logarithmic empirical histogram of h_{agl} of CKA-2020 in yellow and TOPO in blue.

due to lacking in-situ data. Therefore, we test whether in addition to the reconstruction performance the SIF prediction and reflectance estimation are independent of the h_{agl} variation. Fig. 8 (a) shows that the SIF predictions of m_{TOPO} and m_{CKA} have a constant mean over most of the covered height range. This is to be expected in the case of a homogeneous distribution of fluorescence emitting surfaces. The decoupling of h_{agl} from the SIF prediction is only invalid in the range $h_{\mathrm{agl}} < 0.5$ km where both m_{TOPO} and m_{CKA} have a larger mean SIF prediction than in the rest of the height range. There are however significantly less HyPlant pixels falling in this range such that the homogeneity assumption is weakened due to a decreased statistical relevance.

In Fig. 9 we show an exemplary HyPlant acquisition that highlights the independence of the achieved SIF prediction and the reconstruction performance from h_{agl}. Both SIF and the fractional residual $\Delta L/L_{\mathrm{HyP}}$ are unaffected by the topographic variation over the hill slope in the image center. The SIF predictions differ only slightly in m_{TOPO} and m_{CKA} due to differences in the finetuning training data set.

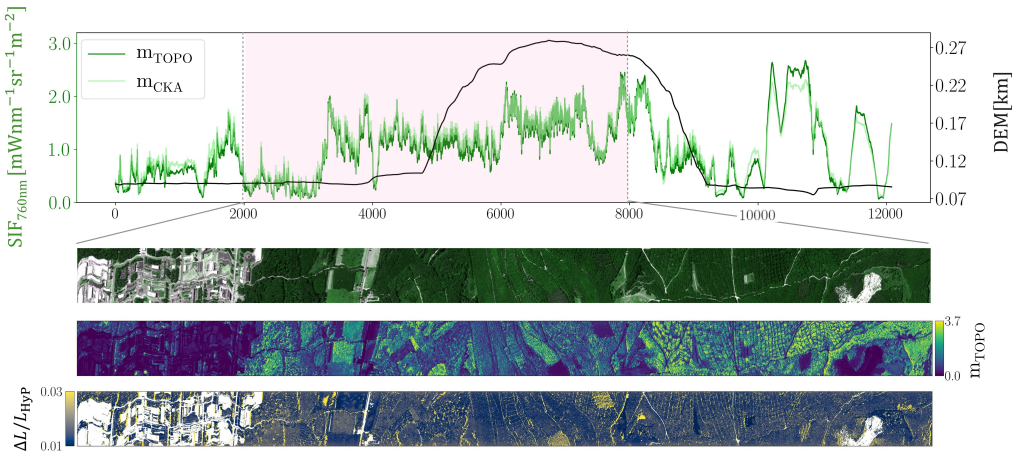


Figure 9: Top row: Shown are the SIF prediction of TOPO(1) and CKA(1) along the ground surface altitude h_{gnd} derived from a matching Digital Elevation Map (DEM). Lower rows: shown are a false color image of the HyPlant at-sensor radiance, the pixelwise SIF prediction of TOPO and the relative reconstruction error of TOPO in the spectral window $\mathcal{W}$ in a subset of the HyPlant acquisition displayed in the top row.

Concluding, we can observe that the influence h_{agl} on EmSFMNN's reconstruction performance, SIF and reflectance prediction is small. The variation of reconstruction errors, SIF and reflectance of both m_{TOPO} and m_{CKA} do not vary systematically with h_{agl}. It could be observed that the choice of the training data set for finetuning had a larger effect on EmSFMNN's reconstruction performance than the topographic variation indicating that EmSFMNN can compensate for the variability in the atmospheric transfer with the specific choice of RTM emulation adopted in this contribution.

5. Discussion

5.1. Simulation and Emulator Design and Limitations

In this work the integrated use of a polynomial emulator and self-supervised neural network training could be shown to yield both fast and accurate estimation of SIF at 760 nm. Two design choices of the simulation model are of particular relevance to the discussion of EmSFMNN's performance.

Firstly, the model underlying the simulation tool and the emulator was set up to not feature any cross-correlation between input parameters. Making use of a physiologically plausible model such as SCOPE [7] relation-

ships would have confounded EmSFMNN's capacity to fit the at-sensor radiance signal purely on the basis of physical principles with possible cross-correlations between estimated parameters. As an example, it is well known that both the reflectance in the spectral region of photosynthetically active radiation (PAR) and the total fluorescence emission amplitude F_{737} are strongly correlated in green vegetation due to a common dependency on leaf chlorophyll concentration [40]. Reconstructed SIF products involving spaceborne reflectance products such as the MODIS-based RSIF [41] and RTSIF [42], reconstructing the TROPOMI SIF product, make use of this relationship. Detailed studies with field data could also establish cross-correlations in reflectance-based features and top-of-canopy SIF derived from airborne platforms resulting from structural effects [43, 44] and biochemical processes related to non-photochemical quenching [45]. However, such dependencies, if incorporated a-priori in the predictor modelling assumptions, may well induce larger gradients in the self-supervised loss than the small at-sensor fluorescence signal. As a consequence, the influence of such correlations on the feature-based optimization and ultimately on the SIF estimate may trump the physical and causal relationship leveraged in EmSFMNN.

Secondly, this contribution has focused on SIF retrieval in a narrow spectral range. The spectral range around the O_2-A absorption band of the simulation data base has allowed us to (1) parametrize the reflectance and fluorescence with simple functions (second-order polynomial, Gaussian) and (2) use a polynomial model to approximate the simulation data base. The polynomial form chosen for the emulator was advantageous to the EmSFMNN set-up as it allowed an easy integration of the emulator in the neural network architecture: both forward pass and backward gradient computation were achieved by implementing the emulator as a fixed linear layer.

The use of the plain polynomial emulator e_4 [28, 27] for EmSFMNN has led to subpar performance with respect to in-situ FLOX measurements. While the lacking spectrally explicit sensor characterization did not lead necessarily to decreased performance in terms of spectral reconstruction residuals, the emulator model's incompleteness has caused systematic errors in the signal decomposition. As a consequence, we have implemented an emulator capable of simulating HyPlant at-sensor radiances with bandwise spectral shifts with an efficient approximation $\tilde{e}_4$. This approximation could be shown to yield acceptable relative errors peaking at $\sim 3.5\%$ with respect to the exact, but computationally demanding emulator solution e_4^{Λ}. The error

incurred by the approximation in the O_2-A absorption band may, however, still be significant in terms of reconstruction accuracy considering that the mean fluorescence emission at 760 nm in HyPlant acquisition amounts also to $< 3\%$ of the mean at-sensor signal. The validation of the SIF prediction of EmSFMNN models integrating this approximate emulator $\tilde{e}_4$ proved sufficient to outperform all baseline methods in terms of accuracy.

While we have shown the application of EmSFMNN to HyPlant FLUO data in this contribution, the EmSFMNN approach to retrieve SIF can be implemented for data from different imaging sensors (e.g., [25]) and spectral regions. Since the network architecture interacts with the data in the loss only through the interface of the emulator representing the physical constraints of the retrieval problem, such a change in the data modality would simply necessitate adapting the emulator. In particular, the modelling of the sensor in the simulation tool [11] and an extension of the reflectance and fluorescence parametric functions to the new spectral range would be required. Further research in emulator representations of simulated hyperspectral at-sensor radiance including bandwise sensor characterization is thus warranted. While the simple polynomial approach adopted here was suitable for the spectral range and simulation model that had been fixed for the EmSFMNN O_2-A SIF retrieval problem on HyPlant data, the integration of different emulator architectures [46, 47, 48] may become necessary for retrieval in different data modalities.

5.2. Prediction of Atmospheric Variables

We have tested the quality of EmSFMNN's signal decomposition with respect to its reconstruction performance and the agreement of its SIF prediction with in-situ measurements. We could not validate the accuracy of the predicted atmospheric variables (water vapour content H_2O and aerosol optical thickness AOT_{550}) with direct measurements. The prediction of these variables is understood to be very challenging in the setting adopted in the presented retrieval method since (i) the sensitivity of the at-sensor radiance to water vapour and AOT_{550} in the fitting spectral window (750 - 770 nm) is small [11], (ii) the variation of both parameters in the training data is expected to be small due to similar meteorological conditions during HyPlant campaigns. Furthermore, there may be remaining representation insufficiencies of the emulator e_4 that can result in EmSFMNN predictors leveraging the degrees of freedom in these parameters to adjust the atmospheric estimate

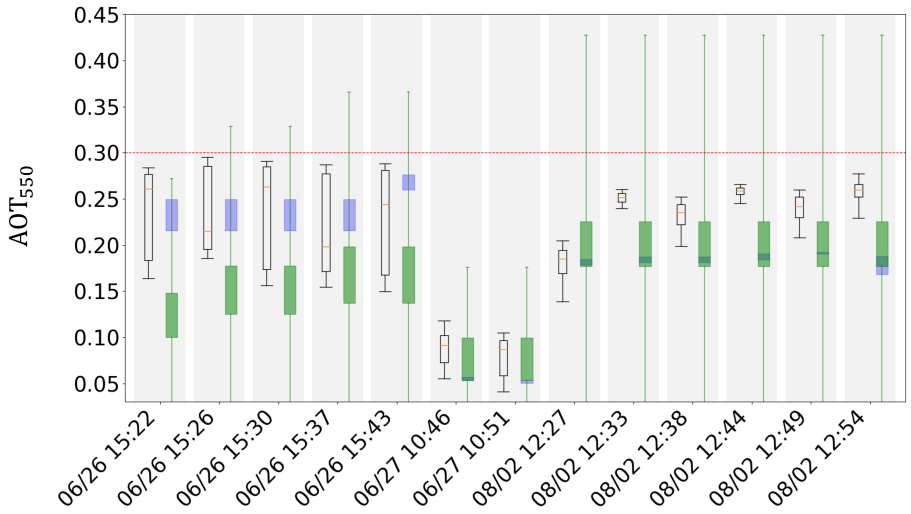


Figure 10: Comparison of AOT_{550} estimates. *Black*: distribution of EmSFMNN AOT_{550} estimates in single HyPlant acquisitions, box width show 25-75 % percentiles and whiskers show 5 -95% percentiles, the median is reported in orange, *Blue*: CIMEL measurements of the JOYCE AERONET station [49] located at $\sim$ 5 km from the SEL-2018 HyPlant acquisitions (left), box width shows the standard deviation of all measurements recorded within 20 minutes of the HyPlant acquisition time, *Green*: Deep Blue AOT_{550} products of MODIS Terra [50] and Aqua [51] at 10 km resolution, box width shows the standard deviation of available Terra and Aqua products within two hours of the HyPlant acquisition time and the whiskers represent the mean of the provided uncertainty of the estimates within a 30 km window around the location of the CKA-2020 estimates. *Red*: maximum AOT_{550} covered in simulation data base.

to the observational data. In particular, we highlight that the at-sensor radiance simulations all have used a standardized atmospheric pressure profile (MODTRAN *mid-latitude summer*) while we have not adapted the emulator to the meteorological conditions at acquisition time. Thus, while H_2O and AOT_{550} were included explicitly in the simulations and EmSFMNN addresses these parameters with a spatial constraint, accurate retrieval of these parameters can not be expected.

We show in Fig. 10, however, that the distributions of AOT_{550} estimates of single HyPlant acquisitions is approximately consistent with AOT_{550} measurements of a CIMEL instrument located in the JOYCE AERONET station [49] nearby the geographical center of HyPlant acquisitions in the SEL-2018 data set ($\sim$ 5 km). We could gather for this analysis CIMEL AOT_{550} measurements with a maximum time difference to the HyPlant acquisition

time of 20 minutes. In particular, we find a strong decrease in AOT_{550} on 2018/06/27 which is reflected in EmSFMNN estimates as well. Furthermore, MODIS Terra and Aqua AOT_{550} estimates from data with a maximum two hour time difference to the HyPlant acquisition are similar to EmSFMNN. The MODIS estimates exhibit large uncertainties, however, such that they must be considered to gauge only very roughly the accuracy of EmSFMNN AOT_{550}.

A more detailed study of parameters pertaining to the atmospheric composition at acquisition time would be of relevance for EmSFMNN if it were to be applied to more susceptible spectral regions covered by full-spectrum retrieval. Furthermore, more extensive analysis could establish the performance benefit of including atmospheric estimates from different sensors as in [25]. Since such a procedure is planned with the FLEX/Sentinel-3 tandem orbit configuration [22], such analysis is relevant especially for further work on the application of EmSFMNN on FLEX data.

5.3. Feature Generalization of EmSFMNN

The validation analyses have shown good performance of EmSFMNN models on data for which the models were not finetuned. In Section 4.2 we could show that $m_{\mathrm{CKA}}(\tilde{\mathrm{e}}_4)$ applied to a HyPlant acquisition from SEL-2018 yielded a mean reconstruction performance of $\epsilon < 3.15\%$ as compared to the result of the finetuned $m_{\mathrm{SEL}}(\tilde{\mathrm{e}}_4)$ of $\epsilon < 2.31\%$. In addition to the reconstruction performance being similar, Section 4.3 could establish that the non-finetuned backbone predictor PRE could estimate SIF outperforming the SFM, iFLD and SFMNN baselines in some of the validation data sets. The SEL-2018 data set is similar to CKA-2020 on which $m_{\mathrm{CKA}}(\tilde{\mathrm{e}}_4)$ was finetuned. Its h_{agl} and h_{gnd} ranges are overlapping. Furthermore, both data sets cover predominantly agricultural fields and exhibit only a small fraction of forested areas such that $m_{\mathrm{CKA}}(\tilde{\mathrm{e}}_4)$ and $m_{\mathrm{SEL}}(\tilde{\mathrm{e}}_4)$ are trained with a similar spectral surface composition. However, the data sets were acquired in different years resulting in varying radiometric sensor calibrations associated with a mean uncertainty of 3%. These results indicate that EmSFMNN generalizes well across HyPlant data sets with large similarities.

Furthermore, a generalization capability of EmSFMNN across topographic changes could be established in Section 4.4. We could show that$m_{\mathrm{CKA}}(\tilde{\mathrm{e}}_4)$, which was finetuned on data exhibiting only small topographic variation, had an improved reconstruction performance over an EmSFMNN instance that was finetuned on the full topographic range present in TOPO. We interpret

this finding such that the learned feature space successfully disentangles features that are unrelated to the topographic change and the associated optical path length differences (i.e. the reflectance and SIF emission). This in turn is corroborated by the observation that both the statistical distribution of the predicted SIF emission as well as of the reflectance are constant across the full topographic range.

The possibility to base the inference of SIF in new data on a single generalized SIF retrieval model is an advantage of the feature based optimization of EmSFMNN over other physical SIF retrieval methods for which a repeated pixelwise or campaign-wise optimization has to be conducted. While we have not conducted validation studies on completely new data sources, that were not included in the pretraining or finetuning training data, we could show that the EmSFMNN could be finetuned successfully to a range of HyPlant data sets without complete retraining. If the importance of finetuning on the SIF prediction performance could be better quantified and reduced, the emulator based SIF retrieval method developed here could therefore prove to be a useful contribution to efficient SIF retrieval method for hyperspectral high-throughput imaging sensors where inference speed is critical.

6. Conclusion

In this work, we have applied EmSFMNN, a novel emulation-based SIF retrieval method first presented by [25], to HyPlant FLUO acquisitions. EmSFMNN utilizes feature-based optimization and hyperspectral RTM emulation to disentangle the fluorescence signal from the at-sensor radiance. It has first been introduced in an application with DESIS data [25]. We have proposed an extension to the originally purely polynomial model used for DESIS to represent spectrally explicit CW and FWHM shifts computationally efficiently. This has allowed for the training of EmSFMNN on a significant fraction of the totality of available HyPlant acquisitions.

The direct SIF validation with in-situ SIF estimates derived from FLOX measurements has shown that the accuracy of finetuned EmSFMNN outperforms both SFMNN as well as traditional baseline methods (SFM, iFLD). Importantly, we could also show that a pretrained backbone EmSFMNN predictor generalized well across the considered HyPlant campaigns such that improved EmSFMNN SIF retrievals could be achieved at a smaller computational cost than traditional pixel-wise optimization. The computational

efficiency of this approach is due to the feature-based nature of EmSFMNN that allows a single model to be used for inference without prior finetuning.

Furthermore, in an analysis with HyPlant acquisition with strong topographic variability, we could show that the set-up allows for a generalization of the application domain of SIF retrieval. The possibility to constrain the retrieval by exact topography and geometrical information has allowed the application of EmSFMNN to HyPlant acquisitions with strong topographic variation where prior retrieval algorithms could not be applied in a straightforward fashion.

Finally, we have presented a small comparison of EmSFMNN predicted AOT_{550} with high-fidelity CIMEL AOT_{550} measurements in a single campaign data set consisting of 13 acquisitions giving first insights into the accuracy of the atmospheric characterization estimated by EmSFMNN. We found a consistent variation of predicted AOT_{550} with the measurements which supports the hypothesis that the disentangling of reflectance, fluorescence and atmospheric components as predicted by EmSFMNN is trustworthy. Further work is, however, necessary to assess EmSFMNN's performance in predicting secondary atmospheric components in general observation conditions.

As HyPlant FLUO is the airborne demonstrator for the spaceborne FLORIS sensor, that will be operated onboard ESA's Earth Explorer mission, this work is relevant for further research in computationally efficient SIF retrieval algorithms for data acquired by FLORIS. While [25] have shown how EmSFMNN could be applied to radiance data acquired on a spaceborne platform, in this work we have focused specifically on the requirements of HyPlant FLUO, a sensor comparable to FLORIS. The encouraging results in terms of precision in both DESIS and HyPlant FLUO suggest that EmSFMNN may be successfully applied to FLORIS data as well.

7. Acknowledgements

This work is part of the project "FluoMap" (Impulsfonds-Förderkennzeichen ZT-I-PF-5-12) funded by the Helmholtz Initiative and Networking Fund, Helmholtz AI, Deutsches Zentrum für Luft- und Raumfahrt (DLR) and Forschungszentrum Jülich GmbH (FZJ). The authors gratefully acknowledge computing time on the supercomputer JURECA [52] at Forschungszentrum Jülich under grant no. fluomap-ct. We gratefully acknowledge the financial support by the European Space Agency (ESA) for airborne data acquisition and data analysis in the frame of the FLEXSense campaign (ESA

Contract No. 4000125402/ 18/NL/NA) and the Photoproxy project (ESA contract No. 4000125731/19/NL/LF). Additionally, HyPlant and FLOX data acquisition have partially been funded by the German Federal Ministry of Education and Research within the German-Plant-Phenotyping Network (DPPN) (project identification number: 031A053), the 'Strukturwandel-Projekt Bioökonomie REVIER', which is funded by the German Federal Ministry of Education and Research (project identification number 031B0918A), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2070–390732324 and the project Land surface Interactions with the Atmosphere over the Iberian Semi-arid Environment (LIAISE) funded by the Centre national de la recherche scientifique (CNRS). Open access is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 491111487.

References

[1] A. Berk, P. Conforti, R. Kennett, T. Perkins, F. Hawes, J. van den Bosch, MODTRAN® 6: A major upgrade of the MODTRAN® radiative transfer code, in: 2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2014, pp. 1–4. doi:10.1109/WHISPERS.2014.8077573.

[2] S. Y. Kotchenova, E. F. Vermote, R. Matarrese, Jr. Frank J. Klemm, Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path radiance, Applied Optics 45 (26) (2006) 6762–6774. doi:10.1364/AO.45.006762.

[3] S. Y. Kotchenova, E. F. Vermote, Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part II. Homogeneous Lambertian and anisotropic surfaces, Applied Optics 46 (20) (2007) 4455–4464. doi:10.1364/AO.46.004455.

[4] C. Emde, R. Buras-Schnell, A. Kylling, B. Mayer, J. Gasteiger, U. Hamann, J. Kylling, B. Richter, C. Pause, T. Dowling, L. Bugliaro, The libRadtran software package for radiative transfer calculations (version 2.0.1), Geoscientific Model Development 9 (5) (2016) 1647–1672. doi:10.5194/gmd-9-1647-2016.

[5] S. Jacquemoud, F. Baret, PROSPECT: A model of leaf optical properties spectra, Remote Sensing of Environment 34 (2) (1990) 75–91. doi:10.1016/0034-4257(90)90100-Z.

[6] J.-B. Feret, C. François, G. P. Asner, A. A. Gitelson, R. E. Martin, L. P. Bidel, S. L. Ustin, G. Le Maire, S. Jacquemoud, PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments, Remote Sensing of Environment 112 (6) (2008) 3030–3043. doi:10.1016/j.rse.2008.02.012.

[7] C. Van Der Tol, W. Verhoef, J. Timmermans, A. Verhoef, Z. Su, An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance, Biogeosciences 6 (12) (2009) 3109–3129. doi:10.5194/bg-6-3109-2009.

[8] C. Van Der Tol, W. Verhoef, A. Rosema, A model for chlorophyll fluorescence and photosynthesis at leaf scale, Agricultural and Forest Meteorology 149 (1) (2009) 96–105. doi:10.1016/j.agrformet.2008.07.007.

[9] S. Jacquemoud, W. Verhoef, F. Baret, C. Bacour, P. J. Zarco-Tejada, G. P. Asner, C. François, S. L. Ustin, PROSPECT+SAIL models: A review of use for vegetation characterization, Remote Sensing of Environment 113 (2009) S56–S66. doi:10.1016/j.rse.2008.01.026.

[10] J.-P. Gastellu-Etchegorry, T. Yin, N. Lauret, T. Cajgfinger, T. Gregoire, E. Grau, J.-B. Feret, M. Lopes, J. Guilleux, G. Dedieu, Z. Malenovský, B. Cook, D. Morton, J. Rubio, S. Durrieu, G. Cazanave, E. Martin, T. Ristorcelli, Discrete Anisotropic Radiative Transfer (DART 5) for Modeling Airborne and Satellite Spectroradiometer and LIDAR Acquisitions of Natural and Urban Landscapes, Remote Sensing 7 (2) (2015) 1667–1701. doi:10.3390/rs70201667.

[11] M. Pato, K. Alonso, J. Buffat, S. Auer, E. Carmona, S. Maier, R. Müller, P. Rademske, U. Rascher, H. Scharr, Simulation Framework for Solar-Induced Fluorescence Retrieval and Application to Desis and Hyplant (Feb. 2025). doi:10.2139/ssrn.5148562.

[12] F. Daumard, Y. Goulas, A. Ounis, R. Pedrós, I. Moya, Measurement and Correction of Atmospheric Effects at Different Altitudes for Remote Sensing of Sun-Induced Fluorescence in Oxygen Absorption Bands, IEEE Transactions on Geoscience and Remote Sensing 53 (9) (2015) 5180–5196. doi:10.1109/TGRS.2015.2418992.

[13] S. Cogliati, W. Verhoef, S. Kraft, N. Sabater, L. Alonso, J. Vicent, J. Moreno, M. Drusch, R. Colombo, Retrieval of sun-induced fluorescence using advanced spectral fitting methods, Remote Sensing of Environment 169 (2015) 344–357. doi:10.1016/j.rse.2015.08.022.

[14] N. Sabater, J. Vicent, L. Alonso, J. Verrelst, E. M. Middleton, A. Porcar-Castell, J. Moreno, Compensation of Oxygen Transmittance Effects for Proximal Sensing Retrieval of Canopy–Leaving Sun–Induced Chlorophyll Fluorescence, Remote Sensing 10 (10) (2018) 1551. doi:10.3390/rs10101551.

[15] S. Cogliati, M. Celesti, I. Cesana, F. Miglietta, L. Genesio, T. Julitta, D. Schuettemeyer, M. Drusch, U. Rascher, P. Jurado, R. Colombo, A Spectral Fitting Algorithm to Retrieve the Fluorescence Spectrum from Canopy Radiance, Remote Sensing 11 (16) (2019) 1840. doi:10.3390/rs11161840.

[16] S. Wieneke, H. Ahrends, A. Damm, F. Pinto, A. Stadler, M. Rossini, U. Rascher, Airborne based spectroscopy of red and far-red sun-induced chlorophyll fluorescence: Implications for improved estimates of gross primary productivity, Remote Sensing of Environment 184 (2016) 654–667. doi:10.1016/j.rse.2016.07.025.

[17] W. Verhoef, H. Bach, Simulation of hyperspectral and directional radiance images using coupled biophysical and atmospheric radiative transfer models, Remote Sensing of Environment 87 (1) (2003) 23–41. doi:10.1016/S0034-4257(03)00143-3.

[18] L. Guanter, L. Alonso, L. Gómez-Chova, M. Meroni, R. Preusker, J. Fischer, J. Moreno, Developments for vegetation fluorescence retrieval from spaceborne high-resolution spectrometry in the O2-A and O2-B absorption bands, Journal of Geophysical Research: Atmospheres 115 (D19) (2010). doi:10.1029/2009JD013716.

[19] A. Damm, L. Guanter, V. C. E. Laurent, M. E. Schaepman, A. Schickling, U. Rascher, FLD-based retrieval of sun-induced chlorophyll fluorescence from medium spectral resolution airborne spectroscopy data, Remote Sensing of Environment 147 (2014) 256–266. doi:10.1016/j.rse.2014.03.009.

[20] T. L. Anderson, R. J. Charlson, D. M. Winker, J. A. Ogren, K. Holmén, Mesoscale Variations of Tropospheric Aerosols*, Journal of the Atmospheric Sciences 60 (1) (2003) 119–136. doi:10.1175/1520-0469(2003)060<0119:MVOTA>2.0.CO;2.

[21] D. R. Thompson, B. H. Kahn, P. G. Brodrick, M. D. Lebsock, M. Richardson, R. O. Green, Spectroscopic imaging of sub-kilometer spatial structure in lower-tropospheric water vapor, Atmospheric Measurement Techniques 14 (4) (2021) 2827–2840. doi:10.5194/amt-14-2827-2021.

[22] M. Drusch, J. Moreno, U. Del Bello, R. Franco, Y. Goulas, A. Huth, S. Kraft, E. M. Middleton, F. Miglietta, G. Mohammed, L. Nedbal, U. Rascher, D. Schüttemeyer, W. Verhoef, The FLuorescence EXplorer Mission Concept—ESA's Earth Explorer 8, IEEE Transactions on Geoscience and Remote Sensing 55 (3) (2017) 1273–1284. doi:10.1109/TGRS.2016.2621820.

[23] D. R. Thompson, N. Bohn, P. G. Brodrick, N. Carmon, M. L. Eastwood, R. Eckert, C. G. Fichot, J. P. Harringmeyer, H. M. Nguyen, M. Simard, A. K. Thorpe, Atmospheric Lengthscales for Global VSWIR Imaging Spectroscopy, Journal of Geophysical Research: Biogeosciences 127 (6) (2022) e2021JG006711. doi:10.1029/2021JG006711.

[24] J. Buffat, M. Pato, K. Alonso, S. Auer, E. Carmona, S. Maier, R. Müller, P. Rademske, B. Siegmann, U. Rascher, H. Scharr, A multi-layer perceptron approach for SIF retrieval in the O2-A absorption band from hyperspectral imagery of the HyPlant airborne sensor system, Remote Sensing of Environment 318 (2025) 114596. doi:10.1016/j.rse.2024.114596.

[25] J. Buffat, M. Pato, K. Alonso, S. Auer, E. Carmona, S. Maier, R. Müller, P. Rademske, U. Rascher, H. Scharr, Retrieval of sun-induced plant fluorescence in the O$_2$-A absorption band from DESIS imagery, accepted for publication in the ECCV Workshop Proceedings at CVPPA 2024. (Nov. 2024). arXiv:2411.08925, doi:10.48550/arXiv.2411.08925.

[26] J. Vicent Servera, J. P. Rivera-Caicedo, J. Verrelst, J. Muñoz-Marí, N. Sabater, B. Berthelot, G. Camps-Valls, J. Moreno, Systematic Assessment of MODTRAN Emulators for Atmospheric Correction, IEEE Transactions on Geoscience and Remote Sensing 60 (2022) 1–17. doi:10.1109/TGRS.2021.3071376.

[27] M. Pato, J. Buffat, K. Alonso, S. Auer, E. Carmona, S. Maier, R. Müller, P. Rademske, U. Rascher, H. Scharr, Physics-based Machine Learning Emulator of At-sensor Radiances for Solar-induced Fluorescence Retrieval in the O_2-A Absorption Band, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2024) 1–10doi:10.1109/JSTARS.2024.3457231.

[28] M. Pato, K. Alonso, S. Auer, J. Buffat, E. Carmona, S. Maier, R. Müller, P. Rademske, U. Rascher, H. Scharr, Fast Machine Learning Simula-

tor of At-Sensor Radiances for Solar-Induced Fluorescence Retrieval with DESIS and Hyplant, in: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, 2023, pp. 7563–7566. doi:10.1109/IGARSS52108.2023.10281579.

[29] B. Siegmann, L. Alonso, M. Celesti, S. Cogliati, R. Colombo, A. Damm, S. Douglas, L. Guanter, J. Hanuš, K. Kataja, T. Kraska, M. Matveeva, J. Moreno, O. Muller, M. Pikl, F. Pinto, J. Quirós Vargas, P. Rademske, F. Rodriguez-Morene, N. Sabater, A. Schickling, D. Schüttemeyer, F. Zemek, U. Rascher, The High-Performance Airborne Imaging Spectrometer HyPlant—From Raw Images to Top-of-Canopy Reflectance and Fluorescence Products: Introduction of an Automatized Processing Chain, Remote Sensing 11 (23) (2019) 2760. doi:10.3390/rs11232760.

[30] J. Buffat, U. Rascher, P. Rademske, B. Siegmann, L. V. Junker-Frohn, D. Emin, HyData: HyPlant FLUO at-sensor radiance data packages and FLOX measurements for SIF retrieval method development from selected campaigns of the years 2018 - 2023 (2024). doi:10.26165/JUELICH-DATA/QKJKPW.

[31] European Space Agency, Technical assistance for the deployment of an advanced hyperspectral imaging sensor during FLEX-EU (Sep. 2017). doi:10.5270/ESA-20835d4.

[32] European Space Agency, Technical Assistance for the Deployment of an advanced hyperspectral imaging sensor during SoyFLEX (Oct. 2017). doi:10.5270/ESA-50a3dd4.

[33] European Space Agency, Technical Assistance for the Deployment of an advanced hyperspectral imaging sensor during SoyFLEX2 (Apr. 2018). doi:10.5270/ESA-24b3118.

[34] European Space Agency, Photoproxy: Technical Assistance for the Photosynthetic-Proxy Experiment (Nov. 2019). doi:10.57780/esa-bb0ea39.

[35] Rascher, Uwe, Siegmann, Bastian, Krieger, Vera, Matveeva, Maria, Quiros, Juan, Muller, Onno, Rademske, Patrick, Herrera, David, Baum, Stephani, Miglietta, Franco, Genesio, Lorenzo, Colombo, Roberto,

Celesti, Marco, Tudoroiu, Marin, Cogliati, Sergio, Carotenuto, Federico, Gioli, Banjamino, Genangelli, Andrea, Hanus, Jan, Houlous, Karel, Migliavacca, Mirco, Martini, David, Martin, M Pilar, Carrara, Arnaud, Moreno, Gerardo, Gonzales-Gascon, Rosario, El-Madany, Tarek, Pacheco-Labrador, Javier, Damm, Alexander, Ahmed, Rifat, Gupana, Remika, Paul-Limoges, Eugenie, Reiter, Ilja, Xueref-Remy, Irène, Mevy, Jean-Philippe, Santonja, Mathieu, Bendig, Juliane, Malenkovsky, Zbynek, Julitta, Tommaso, Burkart, Andreas, FLEX Sentinel Tandem Campaign: Technical Assistance for airborne measurements during the FLEX Sentinel Tandem Experiment (Jul. 2021). doi:10.57780/esa-ae7953d.

[36] Rascher, Uwe, Baum, Stephani, Siegmann, Bastian, Buffat, Jim, Burkart, Andreas, Cogliati, Sergio, Colombo, Roberto, Damm, Alexander, Genesio, Lorenzo, Hanus, Jan, Herrera, David, Julitta, Tommaso, Knopf, Oliver, Miglietta, Franco, Muller, Onno, Quiros, Juan, FLEXSense: Technical Assistance for Airborne Measurements during the FLEX Sentinel Tandem Experiment (Mar. 2022). doi:10.57780/esa-84e5bf5.

[37] Rascher, Uwe, Siegmann, Bastian, Baum, Stephani, HyPlant FLEX simulator on SAFIRE ATR42 for LIAISE Experiment (Jul. 2022).

[38] L. Alonso, L. Gomez-Chova, J. Vila-Frances, J. Amoros-Lopez, L. Guanter, J. Calpe, J. Moreno, Improved Fraunhofer Line Discrimination Method for Vegetation Fluorescence Quantification, IEEE Geoscience and Remote Sensing Letters 5 (4) (2008) 620–624. doi:10.1109/LGRS.2008.2001180.

[39] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention Is All You Need (Aug. 2023). arXiv:1706.03762, doi:10.48550/arXiv.1706.03762.

[40] J. Verrelst, J. P. Rivera, C. van der Tol, F. Magnani, G. Mohammed, J. Moreno, Global sensitivity analysis of the SCOPE model: What drives simulated canopy-leaving sun-induced fluorescence?, Remote Sensing of Environment 166 (2015) 8–21. doi:10.1016/j.rse.2015.06.002.

[41] P. Gentine, S. H. Alemohammad, Reconstructed Solar-Induced Fluorescence: A Machine Learning Vegetation Product Based on MODIS

Surface Reflectance to Reproduce GOME-2 Solar-Induced Fluorescence, Geophysical Research Letters 45 (7) (2018) 3136–3146. doi:10.1002/2017GL076294.

[42] X. Chen, Y. Huang, C. Nie, S. Zhang, G. Wang, S. Chen, Z. Chen, A long-term reconstructed TROPOMI solar-induced fluorescence dataset using machine learning algorithms, Scientific Data 9 (1) (2022) 427. doi:10.1038/s41597-022-01520-1.

[43] P. Yang, C. Van Der Tol, Linking canopy scattering of far-red sun-induced chlorophyll fluorescence with reflectance, Remote Sensing of Environment 209 (2018) 456–467. doi:10.1016/j.rse.2018.02.029.

[44] P. Yang, C. van der Tol, W. Verhoef, A. Damm, A. Schickling, T. Kraska, O. Muller, U. Rascher, Using reflectance to explain vegetation biochemical and structural effects on sun-induced chlorophyll fluorescence, Remote Sensing of Environment (2019). doi:10.1016/j.rse.2018.11.039.

[45] F. Pinto, M. Celesti, K. Acebron, G. Alberti, S. Cogliati, R. Colombo, R. Juszczak, S. Matsubara, F. Miglietta, A. Palombo, C. Panigada, S. Pignatti, M. Rossini, K. Sakowska, A. Schickling, D. Schüttemeyer, M. Stróżecki, M. Tudoroiu, U. Rascher, Dynamics of sun-induced chlorophyll fluorescence and reflectance to detect stress-induced variations in canopy photosynthesis, Plant, Cell & Environment 43 (7) (2020) 1637–1654. doi:10.1111/pce.13754.

[46] J. Verrelst, N. Sabater, J. P. Rivera, J. Muñoz-Marí, J. Vicent, G. Camps-Valls, J. Moreno, Emulation of Leaf, Canopy and Atmosphere Radiative Transfer Models for Fast Global Sensitivity Analysis, Remote Sensing 8 (8) (2016) 673. doi:10.3390/rs8080673.

[47] J. Verrelst, J. P. Rivera Caicedo, J. Muñoz-Marí, G. Camps-Valls, J. Moreno, SCOPE-Based Emulators for Fast Generation of Synthetic Canopy Reflectance and Sun-Induced Fluorescence Spectra, Remote Sensing 9 (9) (2017) 927. doi:10.3390/rs9090927.

[48] B. D. Bue, D. R. Thompson, S. Deshpande, M. Eastwood, R. O. Green, V. Natraj, T. Mullen, M. Parente, Neural network radiative transfer

for imaging spectroscopy, Atmospheric Measurement Techniques 12 (4) (2019) 2567–2578. doi:10.5194/amt-12-2567-2019.

[49] AERONET FZJ-JOYCE, AERONET Site Information Database (2024).
URL https://aeronet.gsfc.nasa.gov/new_web/photo_db_v3/FZJ-JOYCE.html

[50] MODIS Atmosphere Science Team, MODIS/Terra Aerosol 5-Min L2 Swath 10km (2017). doi:10.5067/MODIS/MOD04_L2.061.

[51] MODIS Atmosphere Science Team, MYD04_L2 MODIS/Aqua Aerosol 5-Min L2 Swath 10km (2017). doi:10.5067/MODIS/MYD04_L2.061.

[52] Jülich Supercomputing Centre, JURECA: Data centric and booster modules implementing the modular supercomputing architecture at Jülich supercomputing centre, Journal of large-scale research facilities 7 (A182) (2021). doi:10.17815/jlsrf-7-182.

Appendix D

Publication IV: Retrieval of sun-induced plant fluorescence in the O_2-A absorption band from DESIS imagery

Manuscript accepted at 9th Computer Vision in Plant Phenotyping and Agriculture at ECCV 2024 for publication in ECCV 2024 Workshop Proceedings; Manuscript published online; DOI: 10.48550/arXiv.2411.08925

Jim Buffat[a], Miguel Pato[b], Kevin Alonso[c], Stefan Auer[b], Emiliano Carmona[b], Stefan Maier[b], Rupert Müller[b], Patrick Rademske[a], Uwe Rascher[a], Hanno Scharr[d]

[a]Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Jülich, Germany
[b]Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany
[c]Starion Group c/o European Space Agency (ESA), Largo Galileo Galilei, Frascati 00044, Italy
[d]Forschungszentrum Jülich GmbH, Institute of Advanced Simulations, IAS-8: Data Analytics and Machine Learning, Jülich, Germany

Corresponding author: Jim Buffat, j.buffat@fz-juelich.de

Author contributions: **Jim Buffat**: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Miguel Pato**: Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization, Data Curation. **Kevin Alonso**: Writing — review & editing, Conceptualization. **Stefan Auer**: Writing – review & editing, Project administration, Funding acquisition, Conceptualization. **Emiliano Carmona**: Writing – review & editing, Conceptualization, Data Curation. **Stefan Maier**: Writing – review & editing, Conceptualization, Data Curation. **Rupert Müller**: Conceptualization. **Patrick Rademske**: Data curation. **Uwe Rascher**: Writing – review & editing, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization. **Hanno Scharr**: Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Overall contribution by Jim Buffat: 80 %

Retrieval of sun-induced plant fluorescence in the O_2-A absorption band from DESIS imagery

Jim Buffat[1], Miguel Pato[2], Kevin Alonso[3], Stefan Auer[2], Emiliano Carmona[2], Stefan Maier[2], Rupert Müller[2], Patrick Rademske[1], Uwe Rascher[1], and Hanno Scharr[4]

[1] Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Jülich, Germany
[2] Remote Sensing Technology Institute, German Aerospace Center (DLR), Oberpfaffenhofen, Germany
[3] Starion Group c/o European Space Agency (ESA), Largo Galileo Galilei, Frascati 00044, Italy
[4] Forschungszentrum Jülich GmbH, Institute of Advanced Simulations, IAS-8: Data Analytics and Machine Learning, Jülich, Germany

Abstract. We provide the first method allowing to retrieve spaceborne SIF maps at 30 m ground resolution with a strong correlation ($r^2 = 0.6$) to high-quality airborne estimates of sun-induced fluorescence (SIF). SIF estimates can provide explanatory information for many tasks related to agricultural management and physiological studies. While SIF products from airborne platforms are accurate and spatially well resolved, the data acquisition of such products remains science-oriented and limited to temporally constrained campaigns. Spaceborne SIF products on the other hand are available globally with often sufficient revisit times. However, the spatial resolution of spaceborne SIF products is too small for agricultural applications. In view of ESA's upcoming FLEX mission we develop a method for SIF retrieval in the O_2-A band of hyperspectral DESIS imagery to provide first insights for spaceborne SIF retrieval at high spatial resolution. To this end, we train a simulation-based self-supervised network with a novel perturbation based regularizer and test performance improvements under additional supervised regularization of atmospheric variable prediction. In a validation study with corresponding HyPlant derived SIF estimates at 740 nm we find that our model reaches a mean absolute difference of 0.78 $\mathrm{mW\,nm^{-1}\,sr^{-1}\,m^{-2}}$.

Keywords: Sun-induced fluorescence · Hyperspectral Sensors · DESIS

1 Introduction

The potential of sun-induced flurorescence (SIF) for agricultural management and phenotyping tasks was recognized early in the development of retrieval algorithms [41]. Since SIF is fuelled by a residual energy flux of photosynthetically active radiation (PAR) that is not consumed by processes related to the plant's photochemistry and thermal energy dissipation it provides a causal link between

radiance measurements and the photosynthetic status of plants [42, 50, 61, 62]. Various studies have utilized this relationship as the theoretical basis for stress detection and monitoring [1, 12, 14, 49, 68], the estimation of photosynthetic activity and gross primary productivity [10, 58, 59, 69], crop monitoring and yield predictions [25, 37, 48, 56] and disease monitoring [8, 51] from SIF estimates derived from remote sensing data at various spatial scales. Quantitative estimates of SIF allow for more sensitive and causally founded physiological assessments compared to purely reflectance based indices commonly used for such tasks. Different studies have shown the increased explanatory power of SIF estimates measured at canopy level in a range of tasks [12, 39, 45, 65].

SIF retrieval methods for a variety of sensors have been developed as the number of airborne and spaceborne sensors with sufficient spectral resolution has increased [43]. However, no spaceborne sensor designed specifically for fluorescence retrieval has yet been operationalized. ESA's Earth Explorer Mission FLEX [16], planned to be launched in 2025, will be the first such instrument. Spaceborne SIF estimates to this day are derived from data acquired by satellite missions for atmospheric characterization (*e.g.*, GOSAT [34], GOME [27, 33], SCIAMACHY [35], OCO-2/3 [17,57], TROPOMI [26,28], TanSAT [67]) as their spectral resolution (SR) and signal-to-noise ratio (SNR) allow for SIF retrieval from Fraunhofer lines [16, 23, 24]. However, the spatial resolution of these atmospheric sensors (> 4 km^2) is insufficient for most agricultural applications. FLEX, on the other hand, will provide radiance data with a pixel size of 300 m which still imposes severe limits on its usability for a wide range of applications in heterogeneous agricultural landscapes.

As an exploratory step towards spaceborne SIF retrieval at high spatial resolution, we therefore propose a deep learning architecture and a loss formulation for the first SIF retrieval from hyperspectral imagery of the DLR Earth Sensing Imaging Spectrometer (DESIS). SIF retrieval from DESIS imagery has the benefit of providing spaceborne SIF products at an unprecedented spatial resolution of 30 m which principally allows for the targeted acquisition of auxiliary validation data at high spatial resolution for the upcoming FLEX mission. However, the SR and SNR of DESIS are insufficient for consistent SIF retrieval with current traditional retrieval methods leveraging data in the O_2-A absorption band [13, 22, 40] where the fluorescence signal contribution to the at-sensor signal has a local maximum. Airborne SIF retrieval with similar methods applied to radiance data at lower SR has however been shown to yield consistent relative SIF estimates [3]. As a solution, we extend the simulation-based self-supervised deep learning approach of [5, 7], called Spectral Fitting Method Neural Network (SFMNN), originally developed with airborne hyperspectral imagery. As in other self-supervised simulation-based learning schemes, this approach leverages the implicit constraints of a differentiable simulator of the physical image generation in the loss [30, 32] and primarily does not rely on labels for training. Further regularization terms that enforce physical and physiological domain constraints allow this encoder-decoder architecture to decompose and reconstruct hyperspectral data in the spectral range around the O_2-A absorption band.

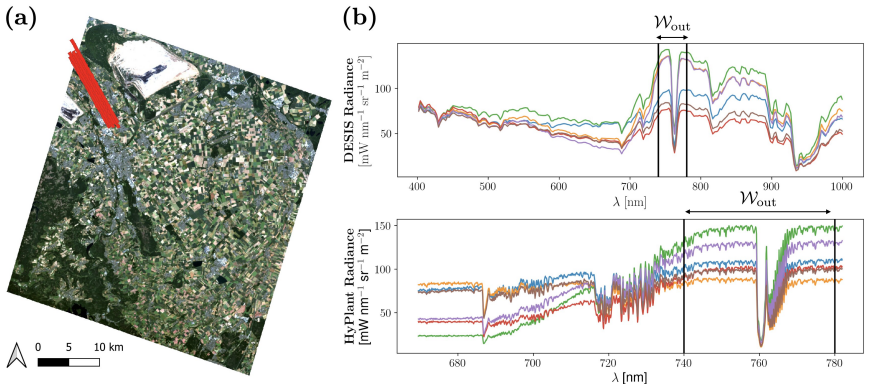


Fig. 1: DESIS and HyPlant data. **(a)** RGB composite of a DESIS acquisition (13/06/2023 14:37 CEST) and, in red, extent of spatially and temporally overlapping HyPlant acquisitions (13/06/2023 14:11 - 14:38 CEST). **(b)** Top: Sample DESIS at-sensor radiance spectra, Bottom: sample HyPlant at-sensor radiance spectra. $\mathcal{W}_{\text{out}}$ denotes the spectral emulator domain.

In this contribution we introduce regularization terms in the SFMNN framework allowing consistent SIF retrieval in DESIS imagery despite its lower SNR and SR. Firstly, we propose a perturbation based augmentation scheme to promote the decorrelation of the predicted SIF from other confounding variables affecting the at-sensor signal. Secondly, we show that including ancillary atmospheric data from DESIS L2A products by means of a secondary supervised downstream learning task improves the performance of our model.

2 Data

2.1 DESIS observation, simulation and emulation

The DLR Earth Sensing Imaging Spectrometer (DESIS) is a hyperspectral imaging sensor onboard the International Space Station (ISS) [38]. It measures at-sensor radiance in 235 bands in the spectral range from 400 nm to 1000 nm with a nominal spectral sampling interval (SSI) of 2.55 nm and a full width at half maximum (FWHM) of 3.55 nm. The spatial dimensions of DESIS acquisitions are fixed to 1024 × 1024 pixels with a nominal pixel width of 30 m. See [2] for a complete description of the DESIS sensor, data products and associated uncertainties. We make use of a polynomial emulator formulation [46, 47] to reconstruct DESIS radiance spectra around the O_2-A absorption band. We found small domain shift errors with respect to smile-corrected L1B DESIS products (see Sec. 4.1). For training, we assembled a data set of 96 DESIS data takes (L1B and L2A in sensor geometry) matching either OCO-3 or HyPlant recordings [6]. The georegistration of DESIS SIF estimates was conducted with the operationally provided DESIS L2A geolayers.

Table 1: Training (*Trn.*) and validation (*Val.*) data sets. N_{px}: number of covered DESIS pixels, N_{acq}: number of DESIS acquisitions (with matching pixels, in the case of data sets *HyPlant* and *OCO-3*). A complete account of the data set is available [6].

Data Set	N_{px}	N_{acq}	Location	Method	Type
HyPlant	10'196 (2020), 18'850 (2023)	4	Jülich (Germany)	[11]	Val.
OCO-3	670	92	Global	[58]	Val.
DESIS	100×10^6	96	Global	–	Trn.

2.2 HyPlant campaigns 2020 and 2023

HyPlant is an airborne spectrometer system providing hyperspectral radiance measurements with an SSI of 0.11 nm and a nominal FWHM of 0.25 nm [55]. It is the airborne demonstrator version for FLEX [16] which is first spaceborne sensor specifically designed for SIF retrieval. As a result, HyPlant measurements have been used for SIF retrieval in yearly field campaigns since 2014 [18–21, 52–54]. For this contribution, spatially overlapping acquisitions of DESIS and HyPlant could be recorded on 13/06/2023. Six HyPlant and two DESIS acquisitions were acquired within small time intervals of 1 - 25 minutes at around 14:30 CEST (cf. Fig. 1 and Tab. 1). Additionally, we found close spatial matches between six HyPlant and two DESIS acquisitions on 23/06/2020 in the same region. In this case the HyPlant acquisitions were recorded at least an hour earlier than the DESIS acquisitions (12:08 CEST). This unique disposition of spatially and temporally matching spaceborne and airborne radiance measurements allowed us to compile a comparative data set of georegistered HyPlant and DESIS SIF estimates. To this end, we processed the HyPlant at-sensor radiance with the Spectral Fitting Method (SFM) [11] to derive high-quality SIF estimates. The alignment of DESIS and HyPlant SIF products involved downscaling (isotropic Gaussian smoothing and spatial resampling) HyPlant SIF to DESIS resolution.

2.3 OCO-3 SIF estimates

OCO-3 is a spectrometer assembly originally designed for the retrieval of column carbon dioxide [17]. As DESIS, OCO-3 is located onboard the ISS. The high SR of the radiance measurements around the O_2-A absorption band of this sensor allows for SIF retrieval in this spectral region similarly to earlier spaceborne sensors designed for the retrieval of atmospheric gas compositions [60]. Since both OCO-3 and DESIS are on the ISS, there exists a set of overlapping acquisitions with small time differences ($<$ 10 minutes). We have identified a set of approximately 100 DESIS acquisitions that are partially covered by OCO-3 measurements, exhibit a low ratio of cloud cover and are flagged to be of acceptable quality. We make use of an OCO-3 SIF product of those acquisitions [15,44] as a complementary performance validation of our DESIS SIF estimates. These OCO-3 SIF estimates were compared to DESIS pixels in a 300 m radius around the center of individual soundings.

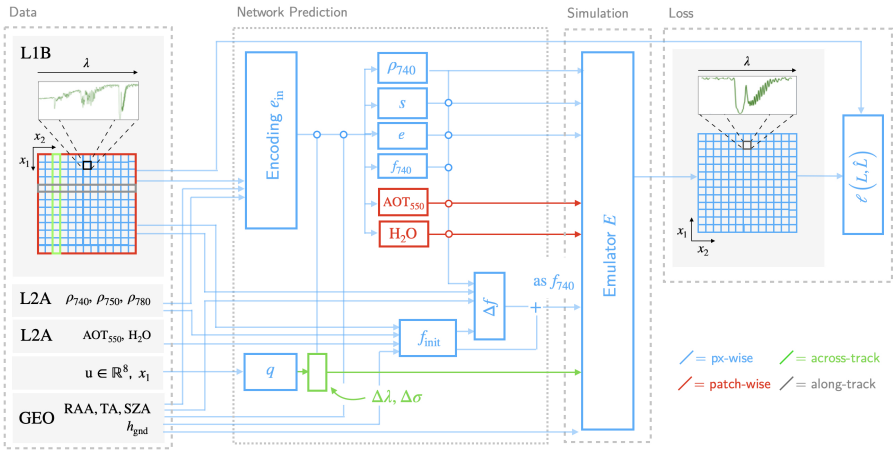


Fig. 2: Proposed network architecture. **Data**: gray blocks denote different data sources: *L1B* smile-corrected DESIS L1B at-sensor radiance, *L2A* reflectance and atmospheric variables provided in the DESIS L2A product, *GEO* geometrical variables from L1C metadata and L2A geolayer: RAA (relative azimuth angle), TA (tilt angle), SZA (sun zenith angle), h_{gnd} (digital elevation model). *other*: u denotes trainable sensor state identifier and x_1 the across-track pixel position. **Network**: variables (ρ_{740}, s, e, f_{740}) predicted by d_{px} and (AOT_{550}, H_2O) predicted by d_{patch} as well as ($\Delta\lambda$, $\Delta\sigma$) predicted by q are passed to the simulation layer implemented as the emulator E [46, 47].

3 Architecture and Simulation-Based Loss

3.1 Architecture

The SIF retrieval method for DESIS imagery presented in this work is based on the Spectral Fitting Method Neural Network (SFMNN) [5, 7]. This network implements in an encoder-decoder type architecture to fit parameters p_j of a simulation model of observational at-sensor radiance data. The simulation model parameterizes the physical signal generation as a function of surface, atmospheric, sensor and geometrical variables. As a result, SIF retrieval is formulated as a feature optimization for optimal spectral decomposition and reconstruction. In order to constrain the solution space, the output dimensionality of the prediction of the simulation parameters is variable (Fig. 2). While surface parameters are allowed to vary in a pixelwise fashion, the atmospheric parameters are constrained to a single scalar value for pixels in a single input patch, i.e., within the same spatial neighbourhood. This is motivated by the large spatial auto-correlation distance of the atmospheric variables which is typically larger than the patches of 30×30 DESIS pixels (900×900 m) used during training. Equally, simulated sensor miscalibration only varies along the across-track dimension.

We implement an encoder, decoders d_{px} and d_{patch} for the surface and atmospheric variables and a module q for the sensor variables. The module architecture consists of stacked multi-layer perceptrons (MLPs) with residual links

(see [7] for a detailed architecture description). The simulation model implemented here exhibits three major differences with respect to [5,7] that allow the application of SFMNN in an adapted form to DESIS data (explained below).

Emulator We replace the approximate simulation model of [5,7] with an emulator of a simulations of DESIS-like radiance spectra around the O_2-A absorption band. To this end, we adopt a simulation tool and emulation set-up introduced by Pato *et al.* [46,47] who show that a polynomial emulator of 4^{th} degree yields an approximation error that is significantly smaller than typical at-sensor fluorescence in a DESIS-like configuration. Polynomial emulators are not widely used for radiance emulation in remote sensing (*e.g.* [63,64]). In our specific case, only the small spectral range $\mathcal{W}_{\text{out}}$ =[740 - 780] nm around the O_2-A band must be covered, however, such that a model with small complexity is able to meet the precision requirements. The polynomial nature of the emulator is advantageous since (i) it can be integrated easily in a feed-forward neural network architecture as a fixed linear layer and (ii) it is computationally efficient such that training and prediction are not significantly affected by it.

Residual Fluorescence Estimation SFMNN is a completely self-supervised approach that does not require any labelled data to be trained to a set of hyperspectral imagery. Preliminary tests with a plain SFMNN approach on DESIS data did not provide SIF estimates with useful sensitivity to the HyPlant and OCO-3 validation data, however. The most likely cause for this is the low SR of DESIS data, especially in comparison to the HyPlant data on which SFMNN was originally developed. We therefore adopt a modification to SFMNN whereby the fluorescence f is not estimated directly from radiance data. Instead, a residual Δf to an initial guess f_{init} with large uncertainty is predicted by a dedicated module from L1C and L2A data (cf. Fig. 2) such that we can interpret

$$f_{740} = f_{\text{init}} + \Delta f, \tag{1}$$

as the model's SIF estimate. A similar approach has notably be implemented by Brodrick *et al.* [4] to improve a low-quality atmospheric radiative transfer model. f_{init} here denotes an initial estimate of the fluorescence emission that we gain from a supervised predictor directly trained on simulated DESIS data. While the simulation-trained model alone yields noisy predictions with subpar performance (cf. Fig. 4), we find that its combination with the self-supervised principles of SFMNN results in significantly improved prediction accuracy.

Sensor state Miscalibration of a hyperspectral sensor can be expressed in terms of changes to a Gaussian approximation of the instrument spectral response function (ISRF) in each spatio-spectral pixel. Commonly, such miscalibrations are parameterized with shifts of its standard deviation $\Delta\sigma$ and shifts of its center wavelength $\Delta\lambda$. Hyperspectral sensors can suffer from changing $\Delta\lambda$ and $\Delta\sigma$ due to mechanical and environmental stresses that change the ISRF and must be

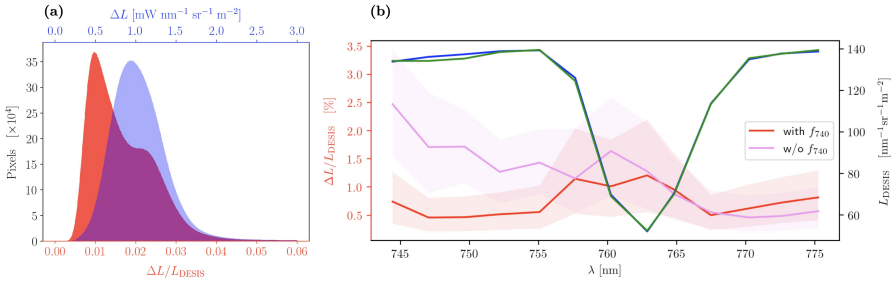


Fig. 3: **(a)** Relative and absolute reconstruction error of best performing model configuration over all DESIS acquisitions. **(b)** *Red/Pink:* Spectrally explicit error distribution in the DESIS acquisition matching the OCO-3 validation data (Fig. 1), light colors denote the 25 - 75 percentiles. *Blue/Green*: Sample reconstruction (blue) of a single spectral DESIS observation (green) matching HyPlant (2023) data.

addressed operationally by periodical recalibration. The issue of mismatching calibrations may be significant due to (i) DESIS' overall low SR and (ii) a subpar smile-correction of L1C in the O_2-A band. In order to alleviate this issue the shifts $\Delta\sigma$ and $\Delta\lambda$ are fitted as a function q of acquisition-specific identifiers u such that sensor drifts, processing changes and artefacts may be accounted for. In order to prevent a too large degree of freedom that could affect the SIF prediction, we implement a module q as a decoder module with only u and sensor position x as inputs such that (i) it is independent of any other input than the identifier u, (ii) multiple acquisitions may have the same identifier (acquisitions of the same date have the same identifier u) and (iii) only across-track variability of the shifts are allowed as is realistic for a push-broom sensor.

Table 2: Dimensions e_{in}, d_{px}, d_{patch}, q and Δf (cf. Fig. 2). Modules consist of stacked MLPs. Each element in *Dim.* denotes the dimension of individual perceptrons in a single MLP, *Reps.* the number of perceptrons in a single MLP (all with the same dimension reported in *Dim.*) and D_p the dropout rate of the output of each MLP. For a detailed architecture description see [7].

Module	Parameters	
Encoder e_{in}	*Dim.*	(1000, 500, 200, 100, 50, 50, 50, 30)
	Reps.	(2, 3, 3, 3, 3, 3, 3, 3)
	D_p	(0.05, 0.01, 0.01, 0.01, 0.005, 0.001)
Decoders d_{px}, d_{patch} and q	*Dim.*	(100, 50, 50, 50)
	Reps.	(3, 2, 2, 2)
	D_p	(0.001, 0.001, 0.0)
Δf	*Dim.*	(1000, 200, 100, 50, 50, 50)
	Reps.	(2, 3, 3, 3, 3, 1)
	D_p	(0.05, 0.005, 0.001)

3.2 Loss formulation

We propose a simulation-based loss $\ell = \ell_{\text{res}} + \ell_m + \ell_{\Delta f} + \ell_N + \ell_c$ where ℓ_{res} evaluates the reconstruction residuals of the model with respect to the observational input, ℓ_m and $\ell_{\Delta f}$ ensure that the network matches prior knowledge, ℓ_N ensures the physiological plausibility of the SIF estimates and ℓ_c denotes a perturbation based regularization that enhances the decorrelation between predicted variables by means of a physically accurate augmentation. Self-supervised learning with radiance observations is addressed by adopting the methodology of [5, 7], where the reconstructed signal is compared to the observation similarly to other self-supervised methods such as masked auto-encoders [29, 31]. A squared residual over the whole spectrum as well as a weighted residual boosting the loss in spectral regions with high average fluorescence contributions punish the network for not reproducing the at-sensor observations. This is implemented by

$$\ell_{\text{res}}\left(L, \hat{L}\right) = \left\langle \left(L - \hat{L}\right)^2 \right\rangle_{\lambda,\, x} + \frac{\gamma_f}{|\mathcal{W}_{\text{out}}|} \left\langle \sum_{\lambda \in \mathcal{W}_{\text{out}}} w_\lambda \left(L(\lambda) - \hat{L}(\lambda)\right)^2 \right\rangle_x , \tag{2}$$

where $\langle \dots \rangle_{\lambda,x}$ denotes the batchwise mean over the spatial and spectral dimension, and where L and $\hat{L}$ denote the observations and emulated predictions in the spectral range $\mathcal{W}_{\text{out}}$. The weighting w_λ is resampled from the weights originally proposed in [5, 7] for the specific sensor characterization of the DESIS sensor. Furthermore, we also adopt the selective gradient backpropagation of the second term which is set to only affect the fluorescence prediction $\hat{f}_{740}$.

The inclusion of prior information on atmospheric variables and the SIF emission are implemented as regularization terms

$$\ell_{\text{m}} = \sum_{k \in \mathcal{K}} \gamma_{p_k} \left(p'_k - \hat{p}_k\right)^2 \quad \text{and} \quad \ell_{\Delta f} = \gamma_{\Delta f} \left(\hat{f} - f_{\text{init}}\right)^2 \tag{3}$$

where f_{init} denotes the prediction of the supervised SIF predictor and where $\mathcal{K} = \{\text{H}_2\text{O},\, \text{AOT}_{550}\}$. We denote by H_2O the water vapour density and by AOT_{550} the aerosol optical thickness at 550 nm, which are derived operationally and distributed with DESIS L2A products. Thus, ℓ_m introduces a supervised regression training of ancillary data from DESIS products as a secondary task. As an alternative strategy, we test in Sec. 4.3 a set-up where the ancillary data is passed directly to the input and the emulator for training and prediction. Accordingly, our model does not provide predictions of the atmospheric parameters (H_2O, AOT_{550}) in this latter variant.

The SIF estimate f_{init}, which we gain from a predictor trained on simulated DESIS data, is included in the loss parallelly to ℓ_m. This effectively controls the range of deviation that the residual module Δf is allowed to introduce. Additionally, the fluorescence estimate of the residual module is controlled by the constraint

$$\ell_N = \gamma_N \, \hat{f} \, \delta\left(\text{NDVI}(L) \leq \tau\right) \tag{4}$$

to ensure vanishing SIF predictions in pixels without vegetation, i.e., with small NDVI [5, 7]. We fix $\tau = 0.15$ in all experiments.

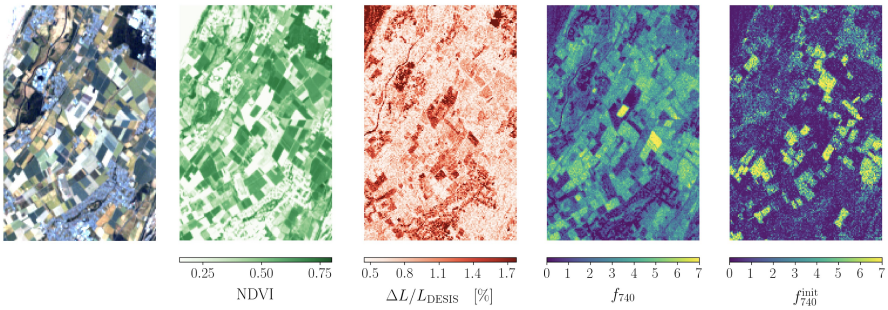


Fig. 4: Overview of an image excerpt of a DESIS acquisition matching HyPlant (2023) validation data (Fig. 3). Left to right: RGB composite, NDVI derived from L2A, relative reconstruction error, fluorescence estimate f_{740}, initial fluorescence estimate f_{740}^{init}.

Finally, we leverage the compact physical description of the generation of at-sensor radiance given by the emulator to derive a perturbation based augmentation for regularization that we will refer to as *consistency regularization*. We denote by g_{p_j} the prediction of emulator input variables p_j. Consequently, g_{p_j} are approximate partial inversions of the emulator e. Perturbations of the input radiance L by δL affect the predictor g_f and vice-versa perturbations δf of the predicted fluorescence $\hat{f}$ affect the emulator:

$$g_f(L+\delta L) = \hat{f} + \delta f \;\text{ and }\; e(\hat{f}+\delta f, p_j) = L_e \;+\delta L_e \tag{5}$$

where the spectral range $\mathcal{W}_{\mathrm{out}}$ of the emulator output L_e is smaller than the range of the observational DESIS spectrum L due to a practical limitation of the emulator design. We can simulate new approximate samples by generating radiance perturbations δL_e. To do this, we write

$$(L+\delta L)(\lambda) \approx L'_{\delta f} = L(\lambda) + \delta L_e \;\delta\left(\lambda \in \mathcal{W}_{\mathrm{out}}\right), \tag{6}$$
$$\text{where }\; \delta L_e = e(\hat{f}+\delta f, p_j) - e(\hat{f}, p_j).$$

Neglecting any simulation errors and inaccuracies due to the perturbation implementation of $L'_{\delta f}$ an optimal solution should yield

$$g_f\left(L'_{\delta f}\right) - \hat{f} = \delta f \;\text{ and }\; g_{p_j}\left(L'_{\delta f}\right) = p_j, \tag{7}$$

since all changes in the perturbed observation can be attributed to a change in fluorescence in this case. We thus formulate the regularization as

$$\ell_c(L, \hat{f}, \hat{p}_j) = \mathbb{E}_{\delta f \sim \mathcal{F}}\left[\left(g_f\left(L'_{\delta f}\right) - (\hat{f}+\delta f)\right)^2 + \left(g_{p_j}\left(L'_{\delta f}\right) - \hat{p}_j\right)^2\right] \tag{8}$$

where $\mathcal{F}$ is the fluorescence range over which e is defined. We implement the expectation as a mean over a uniformly sampled set of δf in each training forward pass.

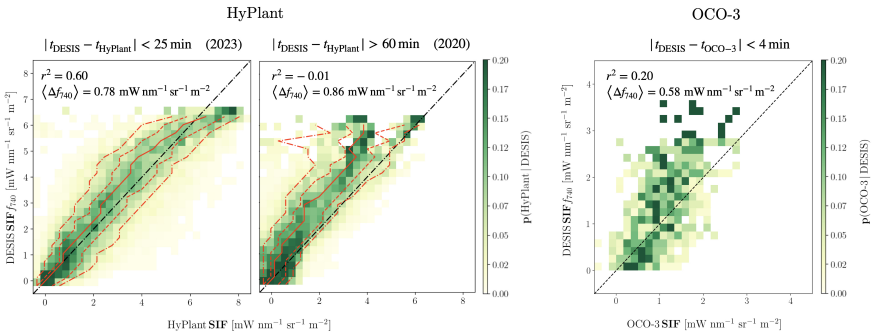


Fig. 5: Conditional 2d-histogram of DESIS SIF estimates of the best performing run ($\gamma_c = 5 \times 10^3$, $\gamma_{\mathrm{AOT}} = 100$) compared to HyPlant (2023), HyPlant (2020) and OCO-3 validation data sets Tab. 1, red dashed lines denote the 10, 25, 75 and 90 percentiles.

4 Results

4.1 Reconstruction Performance & Validation

The network must reconstruct observed spectra with high accuracy. The model trained on all DESIS acquisitions with matching HyPlant or OCO-3 SIF products (Tab. 1) with the proposed loss (with $\gamma_f = 1$, $\gamma_{\Delta f} = 5$, $\gamma_{\mathrm{H_2O}} = 1$ and $\gamma_N = 10$ fixed during preliminary experimental runs and $\gamma_{\mathrm{AOT}} = 100$ and $\gamma_c = 5 \times 10^3$ being the best configuration) reaches a mean relative reconstruction error of 1.6 % (Fig. 3 (a)). The spectral variation of the reconstruction error is small (Fig. 3 (b)) which evidences the model's capacity to plausible signal generation across the spectral domain. We equally observe that the reconstruction error is reduced under inclusion of f_{740} compared to reconstructions where we fixed $f_{740} = 0$.

In Fig. 4 we show an exemplary result. The prediction of f_{740} exhibits a reduced noise level compared to f_{740}^{init} and correlates with the distribution of agricultural fields. In order to assess the performance quantitatively we compare DESIS SIF estimates to matching HyPlant and OCO-3 SIF estimates (Tab. 1). Since existing SIF retrieval methods are not adapted to DESIS data (cf. Sec. 3) we rely on these estimates as ground truth. The same configuration as above yields a mean absolute difference $\langle \Delta f_{740} \rangle_{\mathrm{HyP23}} = 0.78$ mW nm^{-1} sr^{-1} m^{-2} (smaller is better) and a coefficient of determination $r^2_{\mathrm{HyP23}} = 0.6$ (larger is better) in the HyPlant (2023) data set. The DESIS estimates perform worse in a comparison with HyPlant (2020) data ($\langle \Delta f_{740} \rangle_{\mathrm{HyP20}} = 0.86$ mW nm^{-1} sr^{-1} m^{-2} and $r^2_{\mathrm{HyP20}} = -0.01$) due to a large overestimation of our approach. This is expected since the data was recorded closer to solar noon when the diurnal course of SIF peaks [9, 66]. Finally, we find $\langle \Delta f_{740} \rangle_{\mathrm{OCO3}} = 0.58$ mW nm^{-1} sr^{-1} m^{-2} and $r^2_{\mathrm{OCO3}} = 0.2$ compared to OCO-3 data (cf. Fig. 5).

We evaluated the consistency regularization and the inclusion of ancillary data on the SIF prediction. We only validated with respect to the HyPlant (2023) and OCO-3 data sets as the acquisition time difference of HyPlant (2020) would

have introduced large biases. In addition to the metrics above, we report $\langle \Delta f_{740} \rangle$ and r^2 for bias corrected data to differentiate between performance increases due to bias reduction and due to increased explanation of label variance, i.e.,

$$R^2 = r^2 \left(n \circ f_{740}, n \circ \hat{f}_{740} \right) \text{ and } \langle \Delta f_{740} \rangle' = \langle \Delta \left(n \circ f_{740} \right) \rangle \,, \tag{9}$$

where $n(x) = x - \langle x \rangle$. Furthermore, since there is an empirical correlation between SIF and reflectance due to common causal drivers, a stronger validation consists in conditioning the model's performance on reflectance ρ. Subsequently, we define the *reflectance constrained explained variation*

$$\left\langle R^2 \right\rangle_A = |P|^{-1} \sum_{\rho_{780} \in P} R^2(A_{\rho_{780}}), \quad A_{\rho_{780}} = p \left(f_{740} \,\middle|\, \hat{f}_{740},\, |\rho - \rho_{780}| < d\rho \right) \tag{10}$$

where $d\rho = 0.02$ and exclusively reflectance in a single DESIS band (ρ_{780}) is considered. Only the HyPlant (2023) data set is large enough to calculate $\left\langle R^2 \right\rangle_A$ confidently, however. We focus on the reflectance at 780 nm since the influence of SIF on DESIS L2A reflectances at 780 nm is negligible while ρ_{780} also is strongly correlated to vegetation cover. Similarly, we characterize $A_{\rho_{780}}$ by the slope s_A and bias b_A of a linear model fitted to it (cf. Fig. 6 (c)). Since we expect this bias to vanish for perfect predictions we calculate the *mean reflectance dependent bias* $\mathrm{MAE}_b = \langle b_A \rangle$ as an additional performance metric.

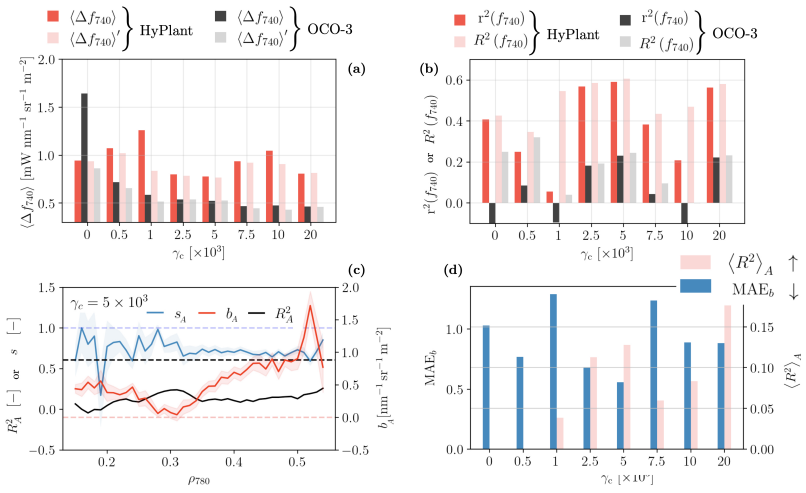


Fig. 6: **(a)** and **(b)**: performance of the model trained with varying γ_c as compared to HyPlant (2023) and OCO-3 SIF. **(c)**: Reflectance constrained metrics for the case $\gamma_{\mathrm{AOT}} = 0$, $\gamma_c = 5 \times 10^3$, light blue and light red colors denote the uncertainty of the least-squares fit to gain s and b, **(d)**: metrics under varying γ_c in HyPlant (2023) data.

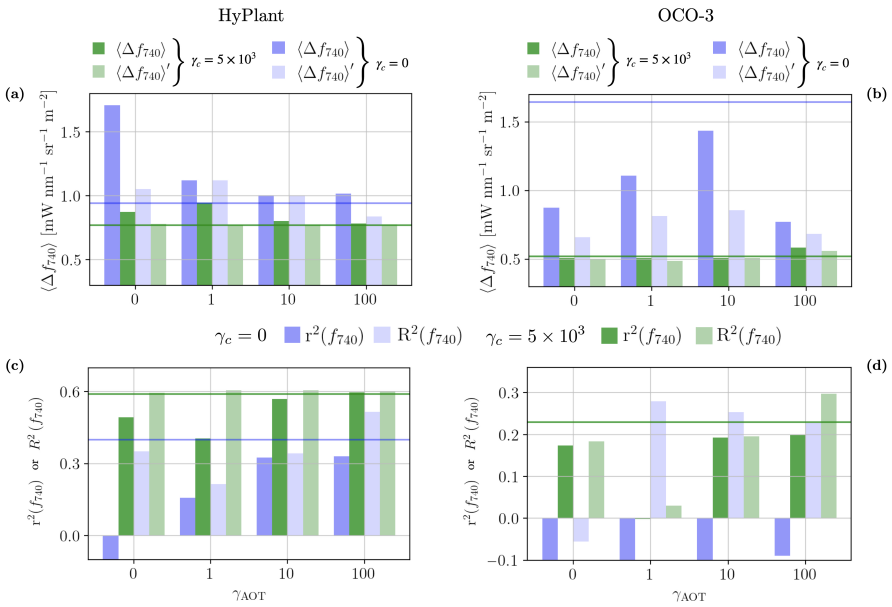


Fig. 7: Model performance with respect to HyPlant (2023) SIF in **(a)** and **(c)** and OCO-3 SIF in **(b)** and **(d)** under variable γ_{AOT}. Light colors indicate $\gamma_c = 0$, dark colors indicate $\gamma_c = 5 \times 10^3$. Horizontal lines indicate the performance of the model runs without ancillary data in Sec. 4.2.

4.2 Consistency Regularization Weighting γ_c

In order to evaluate the impact of the consistency regularization ℓ_c on DESIS SIF, we perform a grid search over γ_c (Fig. 6 (a) and (b)) without including ℓ_m in the training loss. In the OCO-3 data set the inclusion of ℓ_c at all tested weights γ_c outperforms the case $\gamma_c = 0$ in terms of $\langle \Delta f_{740} \rangle$ (smaller is better) and generally also in terms of R^2 (larger is better) (Figs. 6 (a) and (b)). With HyPlant (2023) we find only a localized performance optimum at $\gamma_c \sim 5 \times 10^3$ in terms of $\langle \Delta f_{740} \rangle$ and r^2. We show in Fig. 6 (c) that we find particularly strong overestimation at $\rho_{780} > 0.5$ Improved performance under ℓ_c in HyPlant, however, can be seen in terms of MAE_b and $\langle R^2 \rangle_A$ (Fig. 6 (d)). Specifically, we find $\langle R^2 \rangle_A > 0$ only if the consistency regularization is applied.

4.3 Inclusion of Ancillary Data

To assess the impact of including ancillary data in the SIF retrieval, we conduct a grid search over γ_{AOT} (while fixing $\gamma_{\mathrm{H_2O}} = 1$). First, we establish the performance difference between using the proposed regularization scheme in Eq. (3) and providing the data directly to the input and the emulator. We denote this configuration by $\gamma_{\mathrm{AOT}} = 0$ in Fig. 7. We find decreased r^2 and

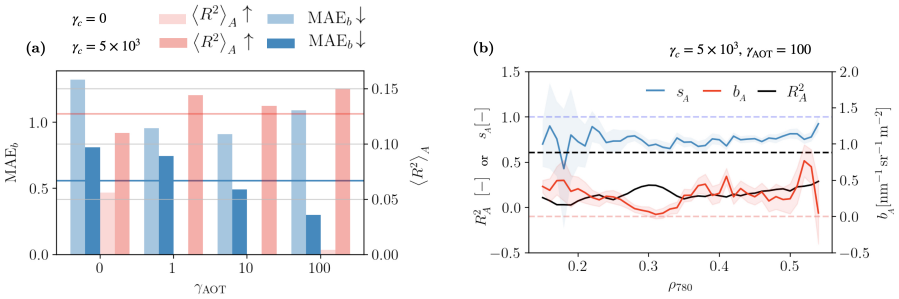


Fig. 8: (a): Reflectance constrained metrics for γ_{AOT}. Light colors indicate model runs with $\gamma_c = 0$, dark colors indicate runs with $\gamma_c = 5 \times 10^3$. Horizontal lines denote the reflectance constrained performance results of the model runs without ancillary data in Fig. 6 (d). **(b):** Reflectance constrained metrics for $\gamma_{\mathrm{AOT}} = 100$, $\gamma_c = 5 \times 10^3$. Light blue and light red denote the uncertainty of the least-squares fit to gain s and b.

$\langle \Delta f_{740} \rangle$ performance in all HyPlant configurations (Fig. 7) compared to cases $\gamma_{\mathrm{AOT}} > 0$. Equally, decreased r^2 performance can be observed in the comparison with OCO-3 SIF estimates. Secondly, we can observe that runs with high γ_{AOT} approximately reach the same SIF prediction performance as model runs without any ancillary data from Sec. 4.2 (Fig. 7 (a) and (c)). Finally, similarly to the results in Sec. 4.2, we can observe a performance increase in the HyPlant (2023) and OCO-3 datasets when using γ_c. This is observable in r^2, $\langle \Delta f_{740} \rangle$, MAE_b and $\langle R^2 \rangle_A$ (Fig. 8 (a)). In particular, the previously observed large overestimation at high ρ_{780} is reduced (Fig. 8 (b)).

5 Discussion

5.1 Decorrelating Impact of Consistency Regularization

The introduction of ℓ_c proved beneficial to the overall SIF retrieval performance. In a study of the performance of $\langle R^2 \rangle_A$, we isolated the model performance independent of the empirical correlation between SIF and ρ_{780}. We found improved performance in terms of $\langle R^2 \rangle_A$ under the inclusion of ℓ_c (Fig. 6 (c) and (d)) indicating that the consistency regularization ℓ_c has the intended effect of decorrelating the target signal from confounding factors. The weighting γ_c has to be chosen carefully, however. Increased MAE_b at large γ_c may have been due (i) to imperfect sample generation in ℓ_c introducing a domain gap between the observations and the augmentations and (ii) a trade-off between reconstruction accuracy and minimization of ℓ_c.

5.2 Inclusion of Ancillary Data Sources

In order to reduce the retrieval problem's ill-posedness we have proposed the use of a regularization that implements the supervised learning of atmospheric emulator prediction variables with ancillary data sources as labels. The regularization

formulates a secondary downstream task in addition to the decomposition of the spectral observations into constituent variables. Importantly, we could see improved SIF retrieval performance with this regularization in terms of a reduced *reflectance dependent bias* MAE_b when used with ℓ_c. As systematic integration of ancillary data is also planned for the FLEX mission by operating it in tandem with Sentinel-3 [16] the retrieval approach explored in this contribution could benefit similar SIF retrieval approaches on FLEX imagery.

6 Conclusion

In this contribution we have presented a deep learning architecture for SIF retrieval from DESIS imagery. This work is the first to use hyperspectral DESIS data for SIF retrieval. A unique data set of spatially and temporally closely matching HyPlant SIF estimates has allowed us to perform a detailed validation study of the methodology proposed in this work. The good performance of our model with respect to these high-quality SIF estimates ($\langle \Delta f_{740} \rangle = 0.78\ \mathrm{mW\,nm^{-1}\,sr^{-1}\,m^{-2}}$, $r^2 = 0.6$) supports our finding that it is possible to derive SIF from DESIS products. Further comparison with a data set of globally distributed OCO-3 SIF estimates could establish the sensitivity of our SIF product in a wider variety of observational and ground conditions and may form the basis for an operational SIF product from DESIS data.

To achieve the good SIF prediction performance, we have extended a self-supervised simulation-based deep learning approach [5, 7]. Several changes to the loss formulation were necessary to address the lower SR and SNR of DESIS imagery. Most importantly, we have (i) introduced a perturbation based augmentation to improve signal decorrelation and (ii) tested the inclusion of ancillary data by formulating a secondary supervised downstream task. We could show that both the perturbation based augmentation and the supervised downstream task formulations improved SIF retrieval performance when comparing both with HyPlant and OCO-3 SIF products. We furthermore could observe improved decorrelation of DESIS SIF from ρ_{780} when making use of the augmentation during training. Since this perturbation based regularization strategy is not restricted to remote sensing data it may be implemented in other simulation-based deep learning applications to decrease the influence of confounding factors.

Acknowledgements

This work is part of the project "FluoMap" (Impulsfonds-Förderkennzeichen ZT-I-PF-5-12) funded by the Helmholtz Initiative and Networking Fund, Helmholtz AI, Deutsches Zentrum für Luft- und Raumfahrt (DLR) and Forschungszentrum Jülich GmbH (FZJ). The authors gratefully acknowledge computing time on the supercomputer JURECA [36] at Forschungszentrum Jülich under grant no. fluomap-ct.

References

1. Ač, A., Malenovský, Z., Olejníčková, J., Gallé, A., Rascher, U., Mohammed, G.: Meta-analysis assessing potential of steady-state chlorophyll fluorescence for remote sensing detection of plant water, temperature and nitrogen stress. Remote Sensing of Environment **168**, 420–436 (Oct 2015). https://doi.org/10.1016/j.rse.2015.07.022
2. Alonso, K., Bachmann, M., Burch, K., Carmona, E., Cerra, D., de los Reyes, R., Dietrich, D., Heiden, U., Hölderlin, A., Ickes, J., Knodt, U., Krutz, D., Lester, H., Müller, R., Pagnutti, M., Reinartz, P., Richter, R., Ryan, R., Sebastian, I., Tegler, M.: Data Products, Quality and Validation of the DLR Earth Sensing Imaging Spectrometer (DESIS). Sensors **19**(20), 4471 (Oct 2019). https://doi.org/10.3390/s19204471
3. Belwalkar, A., Poblete, T., Longmire, A., Hornero, A., Hernandez-Clemente, R., Zarco-Tejada, P.J.: Evaluation of SIF retrievals from narrow-band and sub-nanometer airborne hyperspectral imagers flown in tandem: Modelling and validation in the context of plant phenotyping. Remote Sensing of Environment **273**, 112986 (May 2022). https://doi.org/10.1016/j.rse.2022.112986
4. Brodrick, P.G., Thompson, D.R., Fahlen, J.E., Eastwood, M.L., Sarture, C.M., Lundeen, S.R., Olson-Duvall, W., Carmon, N., Green, R.O.: Generalized radiative transfer emulation for imaging spectroscopy reflectance retrievals. Remote Sensing of Environment **261**, 112476 (Aug 2021). https://doi.org/10.1016/j.rse.2021.112476
5. Buffat, J., Pato, M., Alonso, K., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H.: Deep Learning Based Prediction of Sun-Induced Fluorescence from Hyplant Imagery. In: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium. pp. 2993–2996. IEEE, Pasadena, CA, USA (Jul 2023). https://doi.org/10.1109/IGARSS52108.2023.10282828
6. Buffat, J., Pato, M., Alonso, K., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H.: Selected HyPlant, DESIS and OCO-3 acquisitions for the cross-comparison of sun-induced fluorescence products (2024). https://doi.org/10.26165/JUELICH-DATA/YNYLFQ, https://doi.org/10.26165/JUELICH-DATA/YNYLFQ
7. Buffat, J., Pato, M., Alonso, K., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Siegmann, B., Rascher, U., Scharr, H.: A Multi-Layer Perceptron Approach for SIF Retrieval in the O_2-A Absorption band from Hyperspectral Imagery of the HyPlant Airborne Sensor System (2024), https://doi.org/10.22541/essoar.172411876.66031071/v1, submitted to Remote Sensing of Environment.
8. Calderón, R., Navas-Cortés, J.A., Lucena, C., Zarco-Tejada, P.J.: High-resolution airborne hyperspectral and thermal imagery for early detection of *Verticillium* wilt of olive using fluorescence, temperature and narrow-band spectral indices. Remote Sensing of Environment **139**, 231–245 (Dec 2013). https://doi.org/10.1016/j.rse.2013.07.031
9. Chang, C.Y., Wen, J., Han, J., Kira, O., LeVonne, J., Melkonian, J., Riha, S.J., Skovira, J., Ng, S., Gu, L., Wood, J.D., Näthe, P., Sun, Y.: Unpacking the drivers of diurnal dynamics of sun-induced chlorophyll fluorescence (SIF): Canopy structure, plant physiology, instrument configuration and retrieval methods. Remote Sensing of Environment **265**, 112672 (Nov 2021). https://doi.org/10.1016/j.rse.2021.112672

10. Cheng, Y.B., Middleton, E.M., Zhang, Q., Huemmrich, K.F., Campbell, P.K.E., Corp, L.A., Cook, B.D., Kustas, W.P., Daughtry, C.S.: Integrating Solar Induced Fluorescence and the Photochemical Reflectance Index for Estimating Gross Primary Production in a Cornfield. Remote Sensing **5**(12), 6857–6879 (Dec 2013). https://doi.org/10.3390/rs5126857
11. Cogliati, S., Celesti, M., Cesana, I., Miglietta, F., Genesio, L., Julitta, T., Schuettemeyer, D., Drusch, M., Rascher, U., Jurado, P., Colombo, R.: A Spectral Fitting Algorithm to Retrieve the Fluorescence Spectrum from Canopy Radiance. Remote Sensing **11**(16), 1840 (Jan 2019). https://doi.org/10.3390/rs11161840
12. Damm, A., Cogliati, S., Colombo, R., Fritsche, L., Genangeli, A., Genesio, L., Hanus, J., Peressotti, A., Rademske, P., Rascher, U., Schuettemeyer, D., Siegmann, B., Sturm, J., Miglietta, F.: Response times of remote sensing measured sun-induced chlorophyll fluorescence, surface temperature and vegetation indices to evolving soil water limitation in a crop canopy. Remote Sensing of Environment **273**, 112957 (May 2022). https://doi.org/10.1016/j.rse.2022.112957
13. Damm, A., Guanter, L., Laurent, V.C.E., Schaepman, M.E., Schickling, A., Rascher, U.: FLD-based retrieval of sun-induced chlorophyll fluorescence from medium spectral resolution airborne spectroscopy data. Remote Sensing of Environment **147**, 256–266 (May 2014). https://doi.org/10.1016/j.rse.2014.03.009
14. De Cannière, S., Vereecken, H., Defourny, P., Jonard, F.: Remote Sensing of Instantaneous Drought Stress at Canopy Level Using Sun-Induced Chlorophyll Fluorescence and Canopy Reflectance. Remote Sensing **14**(11), 2642 (Jan 2022). https://doi.org/10.3390/rs14112642
15. Doughty, R., Kurosu, T.P., Parazoo, N., Köhler, P., Wang, Y., Sun, Y., Frankenberg, C.: Global GOSAT, OCO-2, and OCO-3 solar-induced chlorophyll fluorescence datasets. Earth System Science Data **14**(4), 1513–1529 (Apr 2022). https://doi.org/10.5194/essd-14-1513-2022
16. Drusch, M., Moreno, J., Del Bello, U., Franco, R., Goulas, Y., Huth, A., Kraft, S., Middleton, E.M., Miglietta, F., Mohammed, G., Nedbal, L., Rascher, U., Schüttemeyer, D., Verhoef, W.: The FLuorescence EXplorer Mission Concept—ESA's Earth Explorer 8. IEEE Transactions on Geoscience and Remote Sensing **55**(3), 1273–1284 (Mar 2017). https://doi.org/10.1109/TGRS.2016.2621820
17. Eldering, A., Taylor, T.E., O'Dell, C.W., Pavlick, R.: The OCO-3 mission: Measurement objectives and expected performance based on 1 year of simulated data. Atmospheric Measurement Techniques **12**(4), 2341–2370 (Apr 2019). https://doi.org/10.5194/amt-12-2341-2019
18. European Space Agency: Technical assistance for the deployment of an advanced hyperspectral imaging sensor during FLEX-EU (Sep 2017). https://doi.org/10.5270/ESA-20835d4
19. European Space Agency: Technical Assistance for the Deployment of an advanced hyperspectral imaging sensor during SoyFLEX (Oct 2017). https://doi.org/10.5270/ESA-50a3dd4
20. European Space Agency: Technical Assistance for the Deployment of an advanced hyperspectral imaging sensor during SoyFLEX2 (Apr 2018). https://doi.org/10.5270/ESA-24b3118
21. European Space Agency: Photoproxy: Technical Assistance for the Photosynthetic-Proxy Experiment (Nov 2019). https://doi.org/10.57780/esa-bb0ea39
22. Fletcher, K.: Report for Mission Selection. No. 1330/1 in ESA SP, ESA communications production, Noordwijk (2015)

23. Frankenberg, C., Butz, A., Toon, G.C.: Disentangling chlorophyll fluorescence from atmospheric scattering effects in O2 A-band spectra of reflected sun-light. Geophysical Research Letters **38**(3) (2011). https://doi.org/10.1029/2010GL045896
24. Frankenberg, C., O'Dell, C., Guanter, L., McDuffie, J.: Chlorophyll fluorescence remote sensing from space in scattering atmospheres: Implications for its retrieval and interferences with atmospheric CO₂ retrievals. Preprint, Gases/Remote Sensing/Data Processing and Information Retrieval (Mar 2012). https://doi.org/10.5194/amtd-5-2487-2012
25. Guan, K., Berry, J.A., Zhang, Y., Joiner, J., Guanter, L., Badgley, G., Lobell, D.B.: Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence. Global Change Biology **22**(2), 716–726 (Feb 2016). https://doi.org/10.1111/gcb.13136
26. Guanter, L., Aben, I., Tol, P., Krijger, J.M., Hollstein, A., Köhler, P., Damm, A., Joiner, J., Frankenberg, C., Landgraf, J.: Potential of the TROPOspheric Monitoring Instrument (TROPOMI) onboard the Sentinel-5 Precursor for the monitoring of terrestrial chlorophyll fluorescence. Atmospheric Measurement Techniques (2015). https://doi.org/10.5194/amt-8-1337-2015
27. Guanter, L., Zhang, Y., Jung, M., Joiner, J., Voigt, M., Berry, J.A., Frankenberg, C., Huete, A.R., Zarco-Tejada, P., Lee, J.E., Moran, M.S., Ponce-Campos, G., Beer, C., Camps-Valls, G., Buchmann, N., Gianelle, D., Klumpp, K., Cescatti, A., Baker, J.M., Griffis, T.J.: Global and time-resolved monitoring of crop photosynthesis with chlorophyll fluorescence. Proceedings of the National Academy of Sciences **111**(14), E1327–E1333 (Apr 2014). https://doi.org/10.1073/pnas.1320008111
28. Guanter, L., Bacour, C., Schneider, A., Aben, I., Van Kempen, T.A., Maignan, F., Retscher, C., Köhler, P., Frankenberg, C., Joiner, J., Zhang, Y.: The TROPOSIF global sun-induced fluorescence dataset from the Sentinel-5P TROPOMI mission. Earth System Science Data **13**(11), 5423–5440 (Nov 2021). https://doi.org/10.5194/essd-13-5423-2021
29. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked Autoencoders Are Scalable Vision Learners (Dec 2021). https://doi.org/10.48550/arXiv.2111.06377
30. Henderson, P., Ferrari, V.: Learning to Generate and Reconstruct 3D Meshes with only 2D Supervision (Nov 2018)
31. Hong, D., Zhang, B., Li, X., Li, Y., Li, C., Yao, J., Yokoya, N., Li, H., Ghamisi, P., Jia, X., Plaza, A., Gamba, P., Benediktsson, J.A., Chanussot, J.: SpectralGPT: Spectral Remote Sensing Foundation Model. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(8), 5227–5244 (Aug 2024). https://doi.org/10.1109/TPAMI.2024.3362475
32. Jatavallabhula, K.M., Macklin, M., Golemo, F., Voleti, V., Petrini, L., Weiss, M., Considine, B., Parent-Levesque, J., Xie, K., Erleben, K., Paull, L., Shkurti, F., Nowrouzezahrai, D., Fidler, S.: gradSim: Differentiable simulation for system identification and visuomotor control (Apr 2021)
33. Joiner, J., Guanter, L., Lindstrot, R., Voigt, M., Vasilkov, A.P., Middleton, E.M., Huemmrich, K.F., Yoshida, Y., Frankenberg, C.: Global monitoring of terrestrial chlorophyll fluorescence from moderate spectral resolution near-infrared satellite measurements: Methodology, simulations, and application to GOME-2 (Apr 2013). https://doi.org/10.5194/amtd-6-3883-2013
34. Joiner, J., Yoshida, Y., Vasilkov, A.P., Yoshida, Y., Corp, L.A., Middleton, E.M.: First observations of global and seasonal terrestrial chlorophyll fluorescence from space. Biogeosciences **8**(3), 637–651 (Mar 2011). https://doi.org/10.5194/bg-8-637-2011

35. Joiner, J., Yoshida, Y., Guanter, L., Middleton, E.M.: New methods for the retrieval of chlorophyll red fluorescence from hyperspectral satellite instruments: Simulations andapplication to GOME-2 and SCIAMACHY. Atmospheric Measurement Techniques **9**(8), 3939–3967 (Aug 2016). https://doi.org/10.5194/amt-9-3939-2016
36. Jülich Supercomputing Centre: JURECA: Data centric and booster modules implementing the modular supercomputing architecture at jülich supercomputing centre. Journal of large-scale research facilities **7**(A182) (2021). https://doi.org/10.17815/jlsrf-7-182
37. Kira, O., Wen, J., Han, J., McDonald, A.J., Barrett, C.B., Ortiz-Bobea, A., Liu, Y., You, L., Mueller, N.D., Sun, Y.: A scalable crop yield estimation framework based on remote sensing of solar-induced chlorophyll fluorescence (SIF). Environmental Research Letters **19**(4), 044071 (Apr 2024). https://doi.org/10.1088/1748-9326/ad3142
38. Krutz, D., Müller, R., Knodt, U., Günther, B., Walter, I., Sebastian, I., Säuberlich, T., Reulke, R., Carmona, E., Eckardt, A., Venus, H., Fischer, C., Zender, B., Arloth, S., Lieder, M., Neidhardt, M., Grote, U., Schrandt, F., Gelmi, S., Wojtkowiak, A.: The Instrument Design of the DLR Earth Sensing Imaging Spectrometer (DESIS). Sensors **19**(7), 1622 (Apr 2019). https://doi.org/10.3390/s19071622
39. Liu, H., Liu, J., Yin, Y., Walther, S., Ma, X., Zhang, Z., Chen, Y.: Improved Vegetation Photosynthetic Phenology Monitoring in the Northern Ecosystems Using Total Canopy Solar-Induced Chlorophyll Fluorescence Derived From TROPOMI. Journal of Geophysical Research: Biogeosciences **128**(6), e2022JG007369 (2023). https://doi.org/10.1029/2022JG007369
40. Liu, L., Liu, X., Hu, J.: Effects of spectral resolution and SNR on the vegetation solar-induced fluorescence retrieval using FLD-based methods at canopy level. European Journal of Remote Sensing **48**(1), 743–762 (Jan 2015). https://doi.org/10.5721/EuJRS20154841
41. Maier, S.W., Günther, K.P., Stellmes, M.: Sun-Induced Fluorescence: A New Tool for Precision Farming. In: Digital Imaging and Spectral Techniques: Applications to Precision Agriculture and Crop Physiology, chap. 16, pp. 207–222. John Wiley & Sons, Ltd (2004). https://doi.org/10.2134/asaspecpub66.c16
42. Meroni, M., Rossini, M., Guanter, L., Alonso, L., Rascher, U., Colombo, R., Moreno, J.: Remote sensing of solar-induced chlorophyll fluorescence: Review of methods and applications. Remote Sensing of Environment **113**(10), 2037–2051 (Oct 2009). https://doi.org/10.1016/j.rse.2009.05.003
43. Mohammed, G.H., Colombo, R., Middleton, E.M., Rascher, U., van der Tol, C., Nedbal, L., Goulas, Y., Pérez-Priego, O., Damm, A., Meroni, M., Joiner, J., Cogliati, S., Verhoef, W., Malenovský, Z., Gastellu-Etchegorry, J.P., Miller, J.R., Guanter, L., Moreno, J., Moya, I., Berry, J.A., Frankenberg, C., Zarco-Tejada, P.J.: Remote sensing of solar-induced chlorophyll fluorescence (SIF) in vegetation: 50 years of progress. Remote Sensing of Environment **231**, 111177 (Sep 2019). https://doi.org/10.1016/j.rse.2019.04.030
44. OCO Science Team, Gunson, M., Eldering, A.: OCO-3 Level 2 bias-corrected solar-induced fluorescence and other select fields from the IMAP-DOAS algorithm aggregated as daily files, Retrospective processing V10r (2021). https://doi.org/10.5067/NOD1DPPBCXSO
45. Oivukkamäki, J., Atherton, J., Xu, S., Riikonen, A., Zhang, C., Hakala, T., Honkavaara, E., Porcar-Castell, A.: Investigating Foliar Macro- and Micronutrient Variation with Chlorophyll Fluorescence and Reflectance Measurements at the

Leaf and Canopy Scales in Potato. Remote Sensing **15**(10), 2498 (Jan 2023). https://doi.org/10.3390/rs15102498

46. Pato, M., Alonso, K., Auer, S., Buffat, J., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H.: Fast Machine Learning Simulator of At-Sensor Radiances for Solar-Induced Fluorescence Retrieval with DESIS and Hyplant. In: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium. pp. 7563–7566 (Jul 2023). https://doi.org/10.1109/IGARSS52108.2023.10281579
47. Pato, M., Buffat, J., Alonso, K., Auer, S., Carmona, E., Maier, S., Müller, R., Rademske, P., Rascher, U., Scharr, H.: Physics-based Machine Learning Emulator of At-sensor Radiances for Solar-induced Fluorescence Retrieval in the O_2-A Absorption Band. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing pp. 1–10 (2024). https://doi.org/10.1109/JSTARS.2024.3457231
48. Peng, B., Guan, K., Zhou, W., Jiang, C., Frankenberg, C., Sun, Y., He, L., Köhler, P.: Assessing the benefit of satellite-based Solar-Induced Chlorophyll Fluorescence in crop yield prediction. International Journal of Applied Earth Observation and Geoinformation **90**, 102126 (Aug 2020). https://doi.org/10.1016/j.jag.2020.102126
49. Pinto, F., Celesti, M., Acebron, K., Alberti, G., Cogliati, S., Colombo, R., Juszczak, R., Matsubara, S., Miglietta, F., Palombo, A., Panigada, C., Pignatti, S., Rossini, M., Sakowska, K., Schickling, A., Schüttemeyer, D., Stróżecki, M., Tudoroiu, M., Rascher, U.: Dynamics of sun-induced chlorophyll fluorescence and reflectance to detect stress-induced variations in canopy photosynthesis. Plant, Cell & Environment **43**(7), 1637–1654 (Jul 2020). https://doi.org/10.1111/pce.13754
50. Porcar-Castell, A., Tyystjärvi, E., Atherton, J., van der Tol, C., Flexas, J., Pfündel, E.E., Moreno, J., Frankenberg, C., Berry, J.A.: Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: Mechanisms and challenges. Journal of Experimental Botany **65**(15), 4065–4095 (Aug 2014). https://doi.org/10.1093/jxb/eru191
51. Raji, S.N., Subhash, N., Ravi, V., Saravanan, R., Mohanan, C.N., Nita, S., Kumar, T.M.: Detection of mosaic virus disease in cassava plants by sunlight-induced fluorescence imaging: A pilot study for proximal sensing. International Journal of Remote Sensing **36**(11), 2880–2897 (Jun 2015). https://doi.org/10.1080/01431161.2015.1049382
52. Rascher, Uwe, Baum, Stephani, Siegmann, Bastian, Buffat, Jim, Burkart, Andreas, Cogliati, Sergio, Colombo, Roberto, Damm, Alexander, Genesio, Lorenzo, Hanus, Jan, Herrera, David, Julitta, Tommaso, Knopf, Oliver, Miglietta, Franco, Muller, Onno, Quiros, Juan: FLEXSense: Technical Assistance for Airborne Measurements during the FLEX Sentinel Tandem Experiment (Mar 2022). https://doi.org/10.57780/esa-84e5bf5
53. Rascher, Uwe, Siegmann, Bastian, Baum, Stephani: HyPlant FLEX simulator on SAFIRE ATR42 for LIAISE Experiment (Mar 2022)
54. Rascher, Uwe, Siegmann, Bastian, Krieger, Vera, Matveeva, Maria, Quiros, Juan, Muller, Onno, Rademske, Patrick, Herrera, David, Baum, Stephani, Miglietta, Franco, Genesio, Lorenzo, Colombo, Roberto, Celesti, Marco, Tudoroiu, Marin, Cogliati, Sergio, Carotenuto, Federico, Gioli, Banjamino, Genangelli, Andrea, Hanus, Jan, Houlous, Karel, Migliavacca, Mirco, Martini, David, Martin, M Pilar, Carrara, Arnaud, Moreno, Gerardo, Gonzales-Gascon, Rosario, El-Madany, Tarek, Pacheco-Labrador, Javier, Damm, Alexander, Ahmed, Rifat, Gupana, Remika,

Paul-Limoges, Eugenie, Reiter, Ilja, Xueref-Remy, Irène, Mevy, Jean-Philippe, Santonja, Mathieu, Bendig, Juliane, Malenkovsky, Zbynek, Julitta, Tommaso, Burkart, Andreas: FLEX Sentinel Tandem Campaign: Technical Assistance for airborne measurements during the FLEX Sentinel Tandem Experiment (Jul 2021). https://doi.org/10.57780/esa-ae7953d.

55. Siegmann, B., Alonso, L., Celesti, M., Cogliati, S., Colombo, R., Damm, A., Douglas, S., Guanter, L., Hanuš, J., Kataja, K., Kraska, T., Matveeva, M., Moreno, J., Muller, O., Pikl, M., Pinto, F., Quirós Vargas, J., Rademske, P., Rodriguez-Morene, F., Sabater, N., Schickling, A., Schüttemeyer, D., Zemek, F., Rascher, U.: The High-Performance Airborne Imaging Spectrometer HyPlant—From Raw Images to Top-of-Canopy Reflectance and Fluorescence Products: Introduction of an Automatized Processing Chain. Remote Sensing **11**(23), 2760 (Jan 2019). https://doi.org/10.3390/rs11232760
56. Somkuti, P., Bösch, H., Feng, L., Palmer, P.I., Parker, R.J., Quaife, T.: A new space-borne perspective of crop productivity variations over the US Corn Belt. Agricultural and Forest Meteorology **281**, 107826 (Feb 2020). https://doi.org/10.1016/j.agrformet.2019.107826
57. Sun, Y., Frankenberg, C., Wood, J.D., Schimel, D.S., Jung, M., Guanter, L., Drewry, D.T., Verma, M., Porcar-Castell, A., Griffis, T.J., Gu, L., Magney, T.S., Köhler, P., Evans, B., Yuen, K.: OCO-2 advances photosynthesis observation from space via solar-induced chlorophyll fluorescence. Science (2017). https://doi.org/10.1126/science.aam5747
58. Sun, Y., Frankenberg, C., Jung, M., Joiner, J., Guanter, L., Köhler, P., Magney, T.: Overview of Solar-Induced chlorophyll Fluorescence (SIF) from the Orbiting Carbon Observatory-2: Retrieval, cross-mission comparison, and global monitoring for GPP. Remote Sensing of Environment (2018). https://doi.org/10.1016/j.rse.2018.02.016
59. Tagliabue, G., Panigada, C., Dechant, B., Baret, F., Cogliati, S., Colombo, R., Migliavacca, M., Rademske, P., Schickling, A., Schüttemeyer, D., Verrelst, J., Rascher, U., Ryu, Y., Rossini, M.: Exploring the spatial relationship between airborne-derived red and far-red sun-induced fluorescence and process-based GPP estimates in a forest ecosystem. Remote Sensing of Environment **231**, 111272 (Sep 2019). https://doi.org/10.1016/j.rse.2019.111272
60. Taylor, T.E., Eldering, A., Merrelli, A., Kiel, M., Somkuti, P., Cheng, C., Rosenberg, R., Fisher, B., Crisp, D., Basilio, R., Bennett, M., Cervantes, D., Chang, A., Dang, L., Frankenberg, C., Haemmerle, V.R., Keller, G.R., Kurosu, T., Laughner, J.L., Lee, R., Marchetti, Y., Nelson, R.R., O'Dell, C.W., Osterman, G., Pavlick, R., Roehl, C., Schneider, R., Spiers, G., To, C., Wells, C., Wennberg, P.O., Yelamanchili, A., Yu, S.: OCO-3 early mission operations and initial (vEarly) XCO2 and SIF retrievals. Remote Sensing of Environment **251**, 112032 (Dec 2020). https://doi.org/10.1016/j.rse.2020.112032
61. van der Tol, C., Berry, J.A., Campbell, P.K.E., Rascher, U.: Models of fluorescence and photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence. Journal of Geophysical Research: Biogeosciences **119**(12), 2312–2327 (2014). https://doi.org/10.1002/2014JG002713
62. Verrelst, J., Rivera, J.P., van der Tol, C., Magnani, F., Mohammed, G., Moreno, J.: Global sensitivity analysis of the SCOPE model: What drives simulated canopy-leaving sun-induced fluorescence? Remote Sensing of Environment **166**, 8–21 (Sep 2015). https://doi.org/10.1016/j.rse.2015.06.002

63. Verrelst, J., Rivera Caicedo, J.P., Muñoz-Marí, J., Camps-Valls, G., Moreno, J.: SCOPE-Based Emulators for Fast Generation of Synthetic Canopy Reflectance and Sun-Induced Fluorescence Spectra. Remote Sensing **9**(9), 927 (Sep 2017). https://doi.org/10.3390/rs9090927
64. Verrelst, J., Sabater, N., Rivera, J.P., Muñoz-Marí, J., Vicent, J., Camps-Valls, G., Moreno, J.: Emulation of Leaf, Canopy and Atmosphere Radiative Transfer Models for Fast Global Sensitivity Analysis. Remote Sensing **8**(8), 673 (Aug 2016). https://doi.org/10.3390/rs8080673
65. Wang, C., Liu, L., Zhou, Y., Liu, X., Wu, J., Tan, W., Xu, C., Xiong, X.: Comparison between Satellite Derived Solar-Induced Chlorophyll Fluorescence, NDVI and kNDVI in Detecting Water Stress for Dense Vegetation across Southern China. Remote Sensing **16**(10), 1735 (Jan 2024). https://doi.org/10.3390/rs16101735
66. Wang, N., Suomalainen, J., Bartholomeus, H., Kooistra, L., Masiliūnas, D., Clevers, J.G.P.W.: Diurnal variation of sun-induced chlorophyll fluorescence of agricultural crops observed from a point-based spectrometer on a UAV. International Journal of Applied Earth Observation and Geoinformation **96**, 102276 (Apr 2021). https://doi.org/10.1016/j.jag.2020.102276
67. Yao, L., Yang, D., Liu, Y., Wang, J., Liu, L., Du, S., Cai, Z., Lu, N., Lyu, D., Wang, M., Yin, Z., Zheng, Y.: A New Global Solar-induced Chlorophyll Fluorescence (SIF) Data Product from TanSat Measurements. Advances in Atmospheric Sciences **38**(3), 341–345 (Mar 2021). https://doi.org/10.1007/s00376-020-0204-6
68. Zhang, Y., Fang, J., Smith, W.K., Wang, X., Gentine, P., Scott, R.L., Migliavacca, M., Jeong, S., Litvak, M., Zhou, S.: Satellite solar-induced chlorophyll fluorescence tracks physiological drought stress development during 2020 southwest US drought. Global Change Biology **29**(12), 3395–3408 (2023). https://doi.org/10.1111/gcb.16683
69. Zhang, Z., Zhang, Y., Porcar-Castell, A., Joiner, J., Guanter, L., Yang, X., Migliavacca, M., Ju, W., Sun, Z., Chen, S., Martini, D., Zhang, Q., Li, Z., Cleverly, J., Wang, H., Goulas, Y.: Reduction of structural impacts and distinction of photosynthetic pathways in a global estimation of GPP from space-borne solar-induced chlorophyll fluorescence. Remote Sensing of Environment **240**, 111722 (Apr 2020). https://doi.org/10.1016/j.rse.2020.111722