BONNER METEOROLOGISCHE ABHANDLUNGEN

Heft 79 (2017) (ISSN 0006-7156)
Herausgeber: Andreas Hense

Ieda Pscheidt

GENERATING HIGH RESOLUTION PRECIPITATION CONDITIONAL ON RAINFALL OBSERVATIONS AND SATELLITE DATA

BONNER METEOROLOGISCHE ABHANDLUNGEN

Heft 79 (2017) (ISSN 0006-7156)
Herausgeber: Andreas Hense

Ieda Pscheidt

GENERATING HIGH RESOLUTION PRECIPITATION CONDITIONAL ON RAINFALL OBSERVATIONS AND SATELLITE DATA

Generating high resolution precipitation conditional on rainfall observations and satellite data

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Ieda Pscheidt
aus
Sao Bento do Sul, Brasilien

Bonn, März 2017

Diese Arbeit ist die ungekürzte Fassung einer der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn im Jahr 2017 vorgelegten Dissertation von Ieda Pscheidt aus Sao Bento do Sul, Brasilien.

This paper is the unabridged version of a dissertation thesis submitted by Ieda Pscheidt born in Sao Bento do Sul, Brazil to the Faculty of Mathematical and Natural Sciences of the Rheinische Friedrich-Wilhelms-Universität Bonn in 2017.

Anschrift des Verfassers: Address of the author:

Ieda Pscheidt
Meteorologisches Institut der
Universität Bonn
Auf dem Hügel 20
D-53121 Bonn

1. Gutachter: PD Dr. Petra Friederichs, Universität Bonn
2. Gutachter: Prof. Dr. Andreas Hense, Universität Bonn

Tag der Promotion: 06. Juli 2017
Erscheinungsjahr: 2017

Abstract

This study is part of the high resolution reanalysis project proposed for Germany and Europe (Bollmeyer *et al.* 2014) within the framework of the Hans Ertel Centre for Weather Research (HErZ). The reanalysis for Germany assimilates among other variables high resolution rainfall rates. For the most recent years, radar data is assimilated, however, for periods before 2007 this data is not available and another radar-like dataset is required. This study proposes the method HIRAIN to generate an ensemble of probable space-time precipitation fields given a set of observational data. HIRAIN works in two steps. First, a Bayesian statistical model conditional on observations from synoptic stations and on satellite information simulates the latent spatial Gaussian process that drives the occurrence of precipitation exceeding a selected threshold. In a second step, realisations of occurrence/non-occurrence of precipitation exceeding the same thresholds are obtained given the simulated latent process. The occurrence/non-occurrence of precipitation is generated through two different methodologies. HIRAIN is extended to several thresholds of precipitation amount and the final precipitation product is generated from the fields occurrence/non-occurrence of the individual thresholds. A Bayesian approach is used in HIRAIN to provide more realistic fields than those produced by interpolation methods. In the Bayesian approach the data at the observation locations are honored and the spatial covariance structure of the spatial process is reproduced in each realisation. Moreover, the ability to generate ensemble of possible precipitation patterns provides valuable information of precipitation uncertainties that plays also an important role in ensemble reanalysis. HIRAIN produces precipitation dataset with hourly and 4 km resolution. This product presents a more appropriate resolution for the purposes of the reanalysis than the rainfall datasets available by the time the Germany reanalysis project started.

Contents

1. Introduction

Precipitation is an important variable that impacts many sectors of the society. It is part of the hydrological cycle and information of its distribution and intensities helps for predicting and managing natural disasters appropriately as well as for city management, hydrology and agriculture. Many decisions are taken based on the presence or absence of precipitation. Meteorological and hydrological service centers strongly depend on reliable precipitation information among other variables. Good meteorological and hydrological forecasts depend on appropriate data sets that can represent the variability of precipitation. Low spatial and temporal resolution data sets usually lead to poor performance because local precipitation events can not be resolved or the diurnal cycle can not be correctly modeled (Yang, Slingo 2001). Hydrological models have also shown bias when the accuracy of the space-time representation of precipitation is poor (Villarini *et al.* 2009) or when the data is not in scale compatible with that required by the processes being modeled (Vischel *et al.* 2009).

Precipitation information is available from different sources such as direct measurements from rain gauges or from remote-sensing estimation techniques. Rain gauge observations are not able to fully represent the precipitation spatial variability due to the sparseness of the point measurements. Remotely-sensed data as those derived from meteorological radar or satellite estimation techniques present higher spatial resolution and coverage but also larger errors in the estimated amounts. Besides, they are available only for the most recent observational periods.

Given the different limitations of the measurement instruments many approaches have been developed, in order to generate more accurate precipitation products with higher resolution. One approach is to merge precipitation data from different sources. In this scenario, some authors (Adler, Coauthors 2003; De Vera, Terra 2012; Huffman, Coauthors 1997; Janowiak, Xie 1999; Sapiano, Arkin 2009; Vila

et al. 2009; Xiaolan, Lin 2015; Xie, Arkin 1997; Xie *et al.* 2003, e.g.,) merge indirect satellite measurements and direct rain gauge observations. Others merge rain gauge and radar precipitation estimates by employing geostatistical methods, such as ordinary kriging, kriging with external drifts and co-kriging (Bernd *et al.* 2014; Haberlandt 2007; Krajewski 1987) or adjust radar estimates by surface precipitation observations. RADOLAN, which provides a comprehensive high spatio-temporal resolution precipitation dataset for Germany (Bartels *et al.* 2004) is an example of the last method.

Another approach is to simulate precipitation using Bayesian models conditioned on observations and covariates. Within this framework Sanso, Guenni 1999 proposed a model based on transformed truncated multivariate normal distribution to predict monthly precipitation in space and in time. Gaudard *et al.* 1999 used a spatial model to predict mean annual precipitation amounts. Berrocal *et al.* 2008 employed a two-stage model assuming that different Gaussian processes drive precipitation occurrence and precipitation amount. They applied it as a post-processing technique for numerical forecasts on a daily time scale.

In other studies, as the one presented by Clark, Slater 2006, daily precipitation ensembles are modeled with a locally weighted regression model based on rain gauge observations. In their work spatial attributes from station locations are used as covariates to predict the spatial variability in daily totals for a 300 km x 300 km area of western Colorado.

Precipitation is also a fundamental variable for climate change research and monitoring. In order to investigate climate change, researchers need a self-consistent dataset from the past. To this end, many efforts have been made in climate retrospective analysis (reanalysis), which is obtained through the combination of a physical model and observations to produce datasets that describe the past state of the atmosphere, land surface and ocean (European Centre for Medium-Range Weather Forecasts accessed 26 Februar 2016). Since a reanalysis is performed for long periods back in time and precipitation is a very important variable to be assimilated, the availability of historical precipitation datasets is fundamental.

The Hans Ertel Centre for Weather Research (HErZ) (Simmer *et al.* 2015), funded by the German Meteorological Service (Deustcher Wetterdienst - DWD) aimed at producing a high resolution reanalysis based on the numerical weather prediction (NWP) model Consortium for Small Scale Modeling (COSMO). The ob-

jective of the project was to run two high resolution reanalysis, one for Europe and the other with higher resolution for Germany. To this end, COSMO-EU and COSMO-DE that operate with the COSMO model are used (Bollmeyer 2015; Wahl *et al.* 2017). The reanalysis for Germany assimilates among other variables high resolution precipitation rates through a latent heat nudging (LHN) scheme. For this domain the reanalysis was run from 2007 to 2012 and radar precipitation rates were assimilated. However, since radar data is not available before 2007 and in order to assimilate precipitation information in the reanalysis for periods before 2007 another set of radar-like precipitation dataset is required.

After searching for available precipitation products many datasets were found, however, none presented an appropriate time, spatial and temporal resolution. Table 2.2 in chapter 2 presents and discusses in more detalis the available datasets. In view of that, this study proposed a method to generate an appropriate precipitation product for the purposes of the reanalysis.

The proposed method is divided in two steps. First, a Bayesian statistical model conditional on precipitation observations from synoptic stations and on a covariate is used to simulate probabilities of precipitation. In a second step, occurrence/non-occurrence of precipitation above a certain threshold is generated given the simulated probabilities. HIRAIN is extended to several thresholds of precipitation amount and the final precipitation product is generated from the fields occurrence-/non-occurrence of the individual thresholds.

The advantage of a Bayesian approach is that uncertainties of the model parameters are considered for predictions and equal probable scenarios of the spatial random process assumed to generate the observations can be simulated conditioned on the data. The data at the observation locations are honored and the spatial covariance structure of the spatial process is reproduced at each realisation. This leads to more realistic fields than those produced by interpolation methods, which tend to smooth the estimated variable. Moreover, the ability to generate ensemble of possible precipitation patterns can be also very usefull for assimilation in ensemble reanalysis.

Although this research is motivated by the need to produce precipitation data for earlier years than 2007, the fields produced and presented here are for the year 2011, for which the method can be evaluated due to the available radar estimates and a denser network of synoptic stations. The method shown here is applied to

Germany, however the methodology is generally applicable.

The study developed in this doctoral thesis is presented as follows. Chapter 2 provides a brief review of regional reanalysis and presents the HErZ project in more details as well as the importance of this study for the project. Chapter 3 describes the most commonly employed instruments to measure precipitation and presents the datasets used in this work. Chapter 4 reviews important concepts of spatial statistical modeling to provide some background for a better understanding of the proposed method, which is presented in chapter 5. Chapter 6 describes how the method is set up, chapter 7 shows the verification diagnostics, chapters 8 to 11 present the results of this study and lastly, chapter 12 closes this thesis with the conclusions.

2. Regional reanalysis and the Hans Ertel Centre for Weather Research (HErZ)

This chapter concentrates on two main topics. First, a general overview is provided on the foundation of the Hans Ertel Centre for Weather Research (Hans-Ertel-Zentrum für Wetterforschung, HErZ). In the second part, the focus is on a specific research branch, namely the Climate Monitoring and Diagnostics. This branch is responsible for carrying out the regional reanalysis for Europe and Germany and of which this study is part. An overview is given on reanalysis in general and on the regional reanalysis carried out within HErZ. The LHN scheme used to assimilate high resolution precipitation in the reanalysis for Germany is described and the relevance of generating an appropriate precipitation dataset for the purpose of the reanalysis is highlighted.

2.1. HErZ

HErZ was created in 2011 to establish a stronger network between the German universities and atmospheric research centres with the German national weather service DWD (Simmer *et al.* 2015). The foundation of the centre was supported by the German Federal Ministry of Transport, Building and Urban Development and a 12-year core funding period of three four-year phases was foreseen for its development.

The basic research on priority areas that will improve forecasting and climate monitoring at DWD should be carried out at the universities and research centres and then transfered as tools and models to DWD to be employed operationally. On one hand, this network aims at improving the weather service and enabling it to cope with future challenges. On the other hand, it enables the weather service to cooperate and be involved more directly with the environment of universities and

Table 2.1.: Research areas and institutions participating in the first phase of HErZ.

Institutions	Projects
University of Bonn Leibniz Center for Tropospheric Research	Object- based analysis and seamless prediction
Ludwig-Maximilians- University of Munich	Ensemble- based convective scale data assimilation of remote sensing observations
Max Planck Institut for Meteorology - Hamburg	Modeling of convective clouds and stochastic parameterizations
University of Bonn University of Cologne	A high-resolution regional reanalysis
Free University of Berlin	Comunication of weather warnings and extreme weather information for the Berlin conurbation

research centres. There are five areas that were chosen to be the priority areas for research, namely atmospheric dynamics and predictability, data assimilation, model development, climate monitoring and diagnostics, and the optimal use of information from weather forecasting and climate monitoring for the benefit of society. Each of these areas forms a work branch within HErZ. A summarized overview of the branches during the first phase of the project (2011-2014) and the correspondent research focus is presented in table 2.1.

Since this study is part of the first phase of the project carried out within the climate monitoring and diagnostics branch, the following section 2.2 concentrates only on this research group and on its activities.

2.2. The Climate Monitoring and Diagnostics branch

This branch is formed by the Meteorological Institute of the University of Bonn (MIUB) and the Institute for Geophysics and Meteorology of the University of Cologne (IGMK) in collaboration with DWD. The focus of this research group is on continuous development of a self-consistent assessment and analysis of regional climate in Germany and Central Europe from the beginning of the past century to

the present time in an appropriate spatial and temporal resolution. During the first phase of HErZ this branch concentrated on developing a regional reanalysis system and carried it out for Germany and Central Europe.

A reanalysis integrates space and in situ observations through an assimilation scheme method into a physical model. This concept was first proposed by Trenberth, Olson 1988 and Bengtsson, Shukla 1988 to produce estimates of the past climate. The idea of reanalysis can be applied to generate datasets for the atmosphere, the oceans and the biosphere (Bengtsson, Shukla 1988). In atmospheric reanalysis, which is the focus of this research branch, information on meteorological variables in a four-dimensional atmospheric field on a regular grid is produced (Bollmeyer *et al.* 2014; Wahl *et al.* 2017), and the physical model used for that is a NWP model.

Reanalyses have been carried out by different meteorological centres around the world. Available global reanalyses are the National Centers for Environment Predictions/National Center for Atmospheric Research (NCEP/NCAR) global reanalysis (Kalnay *et al.* 1996), the NCEP Climate Forecasting System Reanalysis (CFSR) (Saha *et al.* 2010), the European Centre for Medium - Range Weather Forecasts (ECWMF) ERA-40 (Uppala *et al.* 2005) and ERA-Interim (Dee *et al.* 2011), the National Aeronautics and Space Administration (NASA) Modern-Era Retrospective Analysis for Research and Applications (MERRA) (Rienecker *et al.* 2011) and the Japan Meteorological Agency (JMA) 25-year (Onogi *et al.* 2007) and 55-year (Ebita *et al.* 2011) Reanalysis Project. Current regional reanalysis are the North American Regional Reanalysis (NARR) (Mesinger *et al.* 2006), the European Reanalysis and Observations for Monitoring (EURO4M) (European Reanalysis and Observations for Monitoring Accessed 15 March 2016) and the Artic System Reanalysis (ASR) (Bromwich *et al.* 2010).

For the reanalysis project within the climate monitoring and diagnostics branch different expertise was required for providing the quality controlled and homogenised data set, the statistical post-processing of operational forecasts, the analysis of sytematic model errors of the respective regional model, and the verification and calibration of impact models like e.g. hydrological models. In view of that, the research was carried out in six different working packages (WP), namely: a) WP1, which focused on developing the reanalysis system, b) WP2, responsible for preparing the precipitation dataset for assimilation, c) WP3, in charge of the ver-

ification by satellite and General Observation Period (GOP) data, d) WP4, which performed, analyzed and interpreted reforecasts with the reanalysis system, e) WP5, responsible for provided a comprehensive quality control of the reanalysis datasets and for data publication and f) WP6, in charge of digitisation and homogenization of historical observational data.

The research of WP2 is the topic of this doctoral thesis and the final product of this study will be used for the assimilation scheme in the reanalysis set up in WP1 for Germany. A short overview on WP1 is given first (section 2.2.1) and thereafter the project part of WP2 is motivated (section 2.2.2).

2.2.1. WP1: Regional reanalysis for Europe and Germany

Two reanalysis systems are built using the NWP model COSMO, which is the operational model of DWD with two domains, namely the COSMO-EU with the european domain and COSMO-DE with the domain of Germany. Their horizontal resolutions are 7 km and 2.8 km, respectively. For the reanalysis setup these were changed to 6.2 km and 2 km, and named as COSMO-REA6 and COSMO-REA2, respectively. COSMO-REA6 uses ERA-Interim as lateral boundary condition and COSMO-REA2 uses COSMO-REA6. Observational data are assimilated into COSMO through a nudging technique or Newtonian relaxation, where the prognostic variables of the model are relaxed toward observational data within a predetermined time window (Stauffer, Seaman 1990). The assimilated variables include geopotential, temperature, pressure, humidity and wind data in many levels and from different sources. Rainfall rate is another important variable, however only assimilated in COSMO-REA2 through the LHN scheme (Jones, Macpherson 1997), which is described below. A detailed list with all the assimilated variables, the documentation of the COSMO model as well as a description of the technical implementations to build both reanalysis sytems is provided by Bollmeyer 2015.

2.2.1.1. LHN assimilation scheme in COSMO-REA2

There are different techniques to assimilate data into a physical model. An overview on that is given by Bollmeyer 2015. In COSMO-REA2 a modified version of the LHN scheme used by Jones, Macpherson 1997 is employed. First a brief overview

on the conventional LHN is provided and thereafter the modifications used in COSMO-REA2 are presented.

The conventional LHN scheme is based on the knowledge that latent heat released due to the condensation of water vapor in the clouds is proportional to surface rainfall rate. The precipitation particle takes a certain time interval Δt to travel the distance l from its formation to the ground. Within this time the traveled path can be very complex and the particle can be transported in different directions inside the cloud while it is falling. The model, however, does not consider this complexity and assumes that the particle travels the path l within one single model column and one single time step. With this assumption the surface rainfall rate is proportional to the latent heat released integrated vertically in the column above (Stephan *et al.* 2008), which reads

$$R(l_g) \propto \int_{l_0}^{l_g} \Delta(L_H(l))dl, \tag{2.1}$$

where R is the precipitation rate, $\Delta(L_H)$ the amount of latent heat release, l is the path that a precipitation particle takes from its formation at l_0 to the ground l_g.

The amount of latent heat release $\Delta(L_H)$ changes the temperature in an amount of ΔT given as

$$\Delta T_{LHN}(l) = (\alpha - 1)\frac{1}{cp}\Delta(L_H(l)) \tag{2.2}$$

where

$$\alpha = \frac{R_{obs}}{R_{mod}} \tag{2.3}$$

Jones, Macpherson 1997 used the LHN scheme in a model with 17 km horizontal resolution and a diagnostic computation of precipitation. When precipitation is treated as a prognostic variable as in COSMO-REA2, the conventional LHN scheme needs a few modifications. This is shown by Stephan *et al.* 2008 and is summarized below.

As a prognostic variable, the precipitation takes some time to reach the ground. This is because within a prognostic treatment the particle can be transported into different directions after its formation. The conventional LHN scheme does not

identifies the particle right after its generation and it continues to add or take away energy for some time, even if this is no longer needed. Immediate information on the already initialized precipitation rate is required. This is obtained by introducing a reference precipitation R_{ref} that replaces R_{mod} in equation 2.3. This is given through the vertically averaged precipitation flux,

$$R_{ref} = \frac{1}{z_{top} - z_0} \int_{z_0}^{z_{top}} [\sum_i \rho(z) q_i(z) v_i] dz \tag{2.4}$$

where q_i is the mass fraction of different forms of precipitate, i.e., rain, snow or graupel and v_i is the correspondent velocity of sedimentation. The fluxes are vertically integrated from the layer z_{top} to the ground z_0. The layer z_{top} is defined as the first layer from above, in which the sum of the fluxes is higher than a prespecified threshold (0.1 mm/h).

Other modifications used in COSMO-REA2 include the computation of the LHN increments based only on the vertical layers of the model with a positive latent heat release in order to avoid negative LHN temperature increments and thus cooling where rainfall rate should increase. Due to the fact that a high amount of latent heat is released at the leading edge of convective cells in updraught regions, the precipitation rates (R_{mod}) are low. This together with the high amount of latent heat release leads to very high temperature increments. In order to attenuate that, the limits for the scaling factor α are also changed from 3 to 2 for the upper and from 1/3 to 0.5 for the lower limit. Besides, the scaling factor α is replaced by $ln(\alpha)+1$ to unbias the scheme in terms of adding or taken away absolute amounts of heat energy.

2.2.2. WP2: Generation of precipitation product for the assimilation in COSMO-REA2

The reanalysis for Germany carried out for the most recent period, namely from 2007 to 2012 assimilates radar precipitation rates in COSMO-REA2. However, for periods back in time, meteorological radars are not available and in situ measurements become sparser in time and in space. In order to run the reanalysis for past periods, an appropriate dataset of precipitation is required.

Table 2.2.: Available precipitation datasets.

Dataset	Domain	Period	Time step	Resolution	Data Source, References
CHOMPS	global	1998.01-2007.12	daily	0.25°	satellite, Joseph *et al.* 2009 Yilmaz *et al.* 2010
CMAP	global	1979.01-present	monthly, pentad	2.5°	gauge, satellite, Xie, Arkin 1997
CMORPH	global	2002.12-2016.02	sub-daily, daily	0.25°	satellite, Joyce *et al.* 2004
CPC	global	1979.01-2005.12	daily	0.5°	gauge, Chen *et al.* 2008
CRU TS3.21	global	1901.01-2012.12	climatol., monthly	0.5°	gauge, Harris *et al.* 2014
GPCC	global	1901.01-2010.12	climatol., monthly	0.5° 1° 2.5°	gauge, Becker *et al.* 2013
GPCP	global	1996.01-2015.01	daily	1°	gauge,satellite, Huffman *et al.* 2001
HOAPS	global	1987.01-2008.12	sub-daily, monthly, pentad, climatol.	0.5°	satellite, Andersson *et al.* 2010
PERSIANN CDR	global	1983.01-2015.10	daily, sub-daily	0.25°	gauge, satellite, Ashouri *et al.* 2015
PREC/L	global	1948.01-2013.01	monthly	0.5° 1° 2.5°	gauge,satellite, Chen *et al.* 2002
E-OBS	Europe	1950.01-2015.06	daily	0.25° 0.5°	gauge, Haylock *et al.* 2008
MSG-based gridded	Europe	2005.01-present	15 min	3Km	satellite, Roebeling, Holleman 2009

After searching for precipitation data available before 2007, a large number of datasets were found. However, no one presented the time coverage, the spatial and the temporal resolutions needed for the project purpose. Table 2.2 provides an overview on the datasets available in a global domain and for Germany.

These are gridded datasets and many of them are obtained through data merg-

ing from different sources. One of the highest resolutions is provided by the Meteosat Second Generation (MSG)-based gridded data. This would be appropriate for assimilation, however it is only available since 2005. Other products provide sub-daily data, as the CMORPH, HOAPS and PERSIANN-CDR. Besides the fact that, the temporal resolution is not adequate for the purposes of the project, they only cover the most recent periods. Lastly, the historical precipitation products are daily or monthly data, as for instance CMAP, CRU TS3.21, GPCC, PREC/L, CPC and E-OBS.

Given the lack of appropriate precipitation data, the WP2 proposed the method named HIRAIN to generate an ensemble of probable space-time precipitation fields given a set of observational data. HIRAIN is applied to the most recent period, where radar data and a comprehensive set of rain gauge measurements are available for evaluation. Once an appropriate method is available, ensembles can be generated for periods before 2007.

3. Data sources

The datasets used in this study are provided by rain gauges, meteorological satellite and radar. The following sections give a general overview of these instruments and how precipitation is measured by them. The particular datasets employed in this work are presented in the end of the chapter.

3.0.3. Rain gauges

"A rain gauge is a device used to measure the amount of liquid water obtained from precipitation, including drizzle, rainfall, and the amount of water formed from melting snow and hail" (Strangeways 2007). These devices have been used since mid nineteenth and their number has increased all over the world in time. The first rain gauges have operated manually going through the mechanical and electronical technologies (Strangeways 2007).

The manual rain gauges consist of a graduated cylinder with 2 cm in diameter inside a larger container with diameter of 20 cm and 50 cm tall. The height of the water in the small graduated cylinder is measured, and in case of overflow the excess of water is collected in the larger container and this is poured into another graduated cylinder in order to measure the total rainfall. The gauges are read either daily or when this is not possible weekly or monthly (Strangeways 2007).

The mechanical rain gauges are most used for measuring the average intensity of precipitation and to know when the rain starts and stops. They can be float-operated or weight-operated recordings. The former consists of a rotating drum with a graduate sheet of cardboard and a funnel through which water is collected into a container with a buoy inside. When precipitation falls, the collected water in the container raises the buoy, which raises a pen that is connected to it and to the cardboard through a lever. The pen marks on the paper the precipitation over the time accordingly to the upward or downward movement of the buoy. In the weight-operated the recordings are obtained by weighting the water inside a

suspended container. The weight of the water forces the container downwards and this raises the pen connected to it making it to mark accordingly the precipitation intensity (Strangeways 2007).

The electrical rain gauges or also called the tipping buckets are the most common automatic rain gauges used worldwide nowadays. The measurement resolution can be up to the precipitation amount of 0.1mm. In these rain gauges the collected liquid water is funneled into one of two small cups on either end of a pivoting see-saw beam. When the cup reaches a calibrated weight, the beam drops emptying it and bringing the other cup up into position for the next precipitation increment. The switches of the beam are recorded as "counts" of incremental precipitation. Because the water is not stored in the gauges there is no need for maintenance visits to empty them and since their working mechanism is very simple, there is very little to go wrong. However, they present also some disadvantages. When the buckets tip some water is usually left standing until the next rain. In case it takes a long time for the next rain to occur all the water left standing in the bucket may evaporate (Agrimet Accessed 16 March 2016).

Precipitation measured by rain gauges present also some limitations. Because they are not well distributed over the world, they indicate only the amount in a localized area. In mountain areas and regions with difficult access, the distribution of installed rain gauges is poor, which makes the information of precipitation scarce. The same happens over the oceans, where most of the measurements come from islands due to the difficulty of employing gauges on ships and buoys (Strangeways 2007).

The recordings obtained from rain gauges may contain measurement errors introduced either by the limitations of the instrument itself or by the conditions and environment, where the instrument is located. The catch of a rain gauge may descrease with the increase of wind speed, for instance. Wetting, evaporative losses and snow or ice, that blocks the subsequent rain, may also introduce errors. Changes in instrumentation and changes in the environment surrounding the gauges as vegetation growth and removal, construction and demolition of buildings may influence and introduce a discontinuity into the time series (Groisman, Legates 1994).

3.0.4. Satellite

The state of Earth's atmosphere, surface and ocean has been largely monitored by means of remote sensing by instruments on board of satellites. Satellites do not measure a specific variable itself instead they measure the amount of electromagnetic radiation that is either reflected or emitted by the Earth's surface and atmosphere. Through the amount of electromagnetic energy that reaches the onboard sensors, information on cloud cover, cloud motions, vertical profile of the temperature, rainfall, humidity, ozone or on other aspects as vegetation, dust, fires, among many others can be estimated.

In the Earth atmosphere electromagnetic radiation is unevenly attenuated at different wavelengths of the electromagnetic spectrum. For some spectral regions or bands the atmosphere is transparent, i.e. the radiation is not or it is little absorbed (atmospheric window), while for other frequencies the atmosphere absorbs most of the radiation (absorption window). Based on that, depending on which aspect of the Earth and Earth's atmosphere is intended to be 'measured', the satellite remote sensing works with different spectral regions of the electromagnetic spectrum. The common variables that can be inferred from the measurements through those spectral regions are described next.

The atmospheric window covers the visible and parts of the infrared spectrum, particularly the thermal and mid infrared narrow bands shown as A and B in Figure 3.1. In this window the radiation that reaches the satellite is usually from the surface of the planet or from the top of the clouds. In the absorption window as in the water vapour band only the radiation of the top of the atmosphere can reach the satellite's sensor (Janssen 1993). However, given the fact that an object emits as well as absorbs in this spectral band, vertical profiles of humidity and temperature can be detected.

In the visible spectral band the sensors onboard the satellites are able to detect the solar radiation (Figure 3.1) that reaches the Earth and is reflected back by the oceans, top of the clouds, surface of the Earth, etc. These measurements dependent on the intensity, incidence angle of the sun light reaching a surface and on the surface's albedo. Besides that, it is limited to daytime. This channel does not favour the distinction between low and high clouds, since all clouds have basically the same reflectivity.

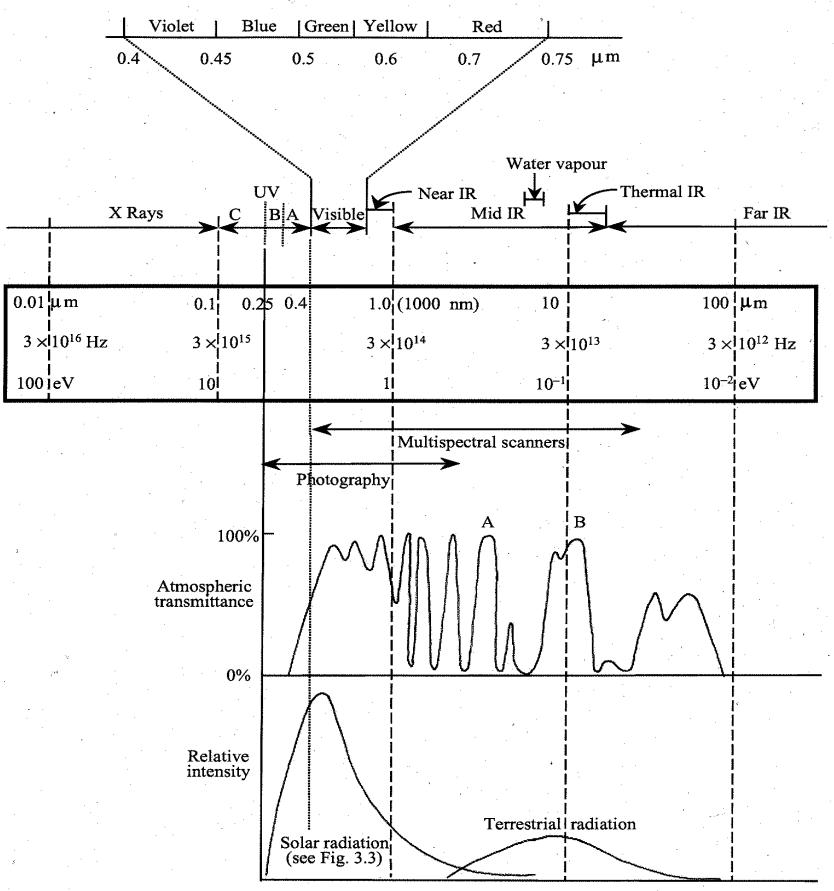

Figure 3.1.: Electromagnetic spectrum, atmospheric transmittance and relative intensity of the solar and terrestrial radiation. (Extracted from Strangeways 2007)

The other limitation is that surface covered with snow present the same reflectivity as clouds and thus it is difficult to differentiate between them in this spectral band.

In the infrared spectral band the radiation is mostly provenient from the Earth radiation and the amount of energy that is detected by satellites provides information about the temperature of the object that emitts it. Figure 3.1 shows that the terrestrial radiation has a maximum at the thermal infrared band indicated through the letter B. Clouds strongly absorb infrared radiation and thus strongly emit it as well. The measurements around the window 11 μm is very useful to detect the temperature of the top of the clouds or of the surface under cloud-free atmosphere (Janssen 1993).

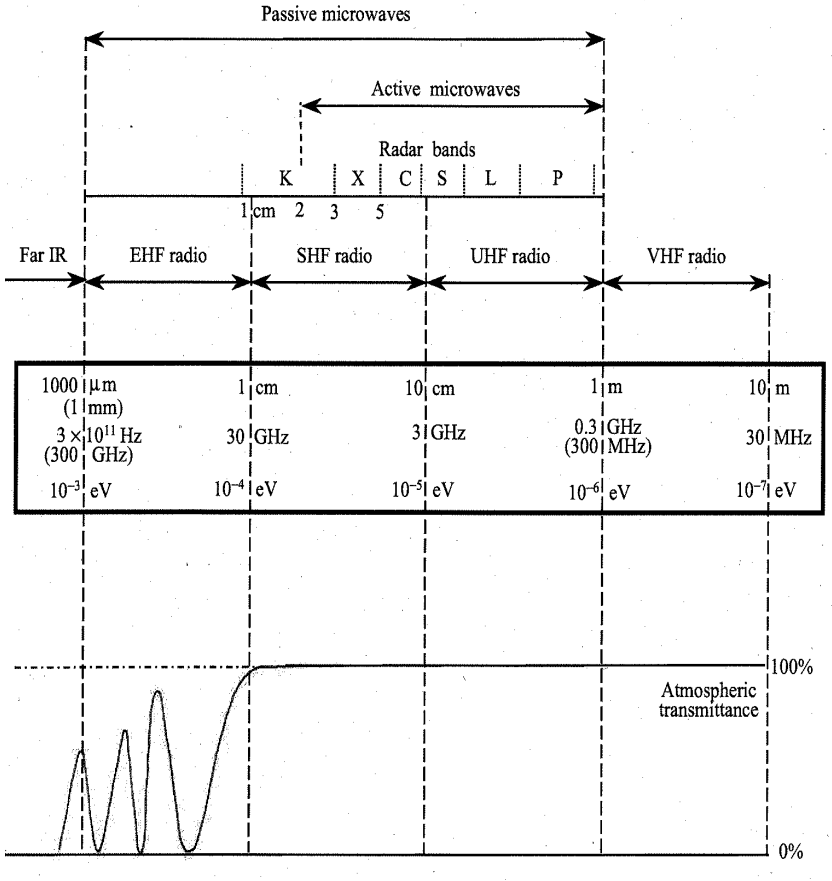

Figure 3.1. (cont.)**:** (Extracted from Strangeways 2007)

Since the temperature in the troposphere decreases with height, the colder the temperature in an infrared image, the higher the cloud top is located. The advantage of this channel in comparison with the visible channel is that these measurements are available all day long and it is possible to distinguish between low and high clouds. However, it is not possible to distinguish between thin and thick clouds. For cloudy sky the radiation of lower clouds can not be detected because the radiation is absorbed in its way from the lower atmosphere to the satellite.

In the microwave wavelengths the atmosphere is transparent for most of the spectral band (Figure 3.1) and the remote sensing depends on the emission, absorption and scattering of the radiation by the surface of the Earth, clouds, ocean, etc (Arkin, Ardanuy 1989). The radiation emitted in the microwave band by the surface of the planet depends on the emissivity and its temperature. A disadvan-

tage of remote sensing in this sprectral band is that emissivity and temperature over land are highly variable and thus the information in the microwave band is also highly variable. This makes the use of the data difficult. Besides, high sensitive sensors are needed to provide reasonable data. However, this spectral band presents the best signal over oceans, since the emissivity is relatively constant (Arkin, Ardanuy 1989). The advantage of microwave over infrared remote sensing is that the radiation is not absorbed by the cirrus clouds and even clouds with water droplets absorb very little in this frequency. This enables the radiation of lower clouds to be detected by the satellite's sensor in the presence of clouds in the higher levels (Janssen 1993).

Satellites can be geostationary or polar orbiting. Geostationary satellites circle the Earth once a day and are approximately 35.880 km distant from the planet. When the satellite orbits the Earth with the same velocity as the planet rotates around itself, it appears to have a fixed position in relation to the planet. When its orbit is located around the equator it is called geosynchronous and in this case the satellite is able to provide a continuous view of the atmosphere from about 70°N to 70°S (Strangeways 2007).

Polar orbiting satellites cover the polar regions, where the geostationary satellites do not have access. They are located at altitudes from about 400 km to 900 km and scan the planet in the north-south orientation. When their orbits are facing the sun they are called sun-synchronous. While the Earth is rotating the satellite scanns a different path in the planet until the whole Earth is scanned (Strangeways 2007). They observe each point on the planet twice a day.

There are many satellites in orbit for climate and weather monitoring. Some of the current geostationary satellites are the GOES-15, GOES-13, GMS, MSG, INSAT-3D, FY-2E and FY-2G. Among the polar-orbiting the NOAA-19, METEOR-3/TOMS, METOP, Landsat, GPM core observatory can be cited (NOAA's Geostationary and Polar-Orbiting Weather Satellites Accessed 8 March 2016).

Since in this study infrared information of MSG network of satellites are used to provide additional information of precipitation occurrence, the remainder of this subsection will focus on this satellites' sensor only.

MSG is a new generation of geostationary satellites developed by the European Space Agency (ESA) in colaboration with the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT). The current MSG network are

the Meteosat- 7, 8, 9, 10 and 11 and they operate over Europe and Africa (Current Satellites Accessed 16 March 2016). They carry onboard the Spinning Enhanced Visible and Infrared Imager (SEVIRI) with 12 spectral channels. SEVIRI provides a new image of the Earth every 15 minutes and the imaging spatial resolution is 3 km at the sub-satellite point for infrared channels and 3 visible channels, whereas 1 km for high resolution visible channels (Schmid Accessed 08 March 2016). For the thermal infrared channels calibration is performed by onboard blackbody targets inserted into the optical path of the instrument.

The channels of the SEVIRI sensor are summarized below (Schmetz *et al.* 2002):

a) Visible channels at 0.6 μm and 0.8 μm are used for cloud detection and tracking, and for monitoring land surface, vegetation and aerosol. b) Near infrared channel at 1.6 μm gives information of aerosol and differentiate between snow and cloud, ice and water clouds. c) Infrared at 3.9 μm provides land and sea surface temperature and is used for detection of low clouds and fog. d) Water vapour channels at 6.2 μm and 7.3 μm detect water vapour and wind in two different levels in the troposphere. e) Infrared at 8.7 μm delivers information of thin cirrus clouds and distinguishes between water and ice clouds. It provides also ozone information. f) Infrared at 10.8 μm and 12.0 μm are used to detect temperature of the top of the clouds. It is also used for detection of volcanic ash clouds and cirrus cloud. g) Infrared at 13.4 μm is the CO_2 absorption channel. In cloud free-areas it provides the temperature of the lower troposphere.

3.0.5. Radar

Radar remote sensing is performed in the microwave band of the electromagnetic spectrum. The radar emits electromagnectic pulses and receives the signal that is scattered back by a target. The signal received back by the radar is called the reflectivity (Z). The time the electromagnetic radiation takes to return to the radar provides information about the location of the target, whereas the signal strength that comes back gives information about the size and distribution of the targets. When moving targets scatter electromagnetic waves there is a shift in the frequency of these waves that is proportional to the velocity of the targets (Rinehart 1997).

Doppler weather radars operate based on this knowledge and deliver informa-

Table 3.1.: Weather radar classification, wavelength and frequency of operation.

Band	Wavelength	Frequency
K	1.7-1.2 cm	18-27 GHz
X	4-2.5 cm	8-12 GHz
C	8-4 cm	4-8 GHz
S	15-8 cm	2-4 GHz
L	30-15 cm	1-2 GHz

tion of the location of the target, the radar reflectivity and the estimated velocity at which the target moves.

Depending on the wavelengths that a radar operates it can be classified as K, X, C, S, or L band radar. Table 3.1 indicates the approximate wavelength and frequencies at which each of them operates.

Radar operating in the lower wavelenghts, as the K and X bands, have small dishes and relatively low costs. Besides, they are easy to transport and may be appropriate when measurements need to be done over different locations. However, the signals are attenuated by heavy rain. Radar operating in the higher wavelengths as the C, S and L bands have the signal less affected by attenuation and are more appropriate to monitor the approaching and development of a precipitating system at large distances from the radar. The disadvantage is the high cost and the large dishes that can reach up to 8 m of diameter, what makes them inappropriate for displacement (Rinehart 1997).

Typically, the radar scans at different elevations. The antenna rotates 360° and does a full scan at the first elevation. After this is completed it tilts up in elevation and performs a full scan. This is performed until the observational volume is fully scanned.

From the radar reflectivity Z, rainfall rate R can be estimated. Marshall *et al.* 1947 were the first to report a relationship between Z and R. This is based on the knowledge that "Z depends on the raindrop size distribution and the size of the drops, and R depends on the raindrop size distribution, the size of the drops and the fall velocity for a given drop diameter" (Fournier 1999). Thus, a $Z-R$ relationship can be build in order to estimate the rainfall rate from the reflectivity.

This is an empirical equation and is given as

$$Z = aR^b, \tag{3.5}$$

where R is in mm/h and Z is in mm^6/m^3. Commonly the logaritmic radar reflectivity is used,

$$Z = 10log_{10}(\frac{z}{1mm^6/m^3}), \tag{3.6}$$

where Z has then units of dBZ, i.e., decibels relative to a reflectivity of 1 mm^6/m^3 (Rinehart 1997).

The paramters a and b in equation 3.5 are estimated and depend on local rainfall conditions and on the radar characteristics. Thus, the parameters can vary from one location to another, among storms, distance from the radar, raindrops distribution, from one radar to another, etc. Usually when not much is known about the weather local conditions the Marshall-Palmer Z-R relationship (Marshall, Palmer 1948) is used, where $a = 200$ and $b = 1.6$. Since this relationship does not always represent the true nature of the rainfall and the radar features, errors may be introduced in the estimates.

Ideally appropriate values for the parametes a and b should be estimated. Methods as the probability matching method (PMM) proposed first by Calheiros, Zawadzki 1987 have been applied to determine appropriate Z-R relationships. In this method radar is calibrated against rain gauge observations, where the probability density function (PDF) of the radar reflectivity is forced to be similar to the PDF of the rain gauge observations. "This approach eliminates the requirement for simultaneous radar-gauge measurements as long as the gauge measurements represent a sample of the local climatology" (Crosson *et al.* 1996).

Large rainfall uncertainties can be obtained by using a Z-R relationship that does not represent the local rainfall distribution and the radar characteristics. Besides, other common sources of errors are the attenuation of the radar beam by heavy rain or clouds, clutter and bright band. Since "much of the rain that falls to the ground begins as ice or snow "(Rinehart 1997) radar may sometimes receive the scattered signal from an ice particle and interpret it as a raindrop. When a snowflake reaches the melting level in the atmosphere, i.e., the 0° isotherm, it

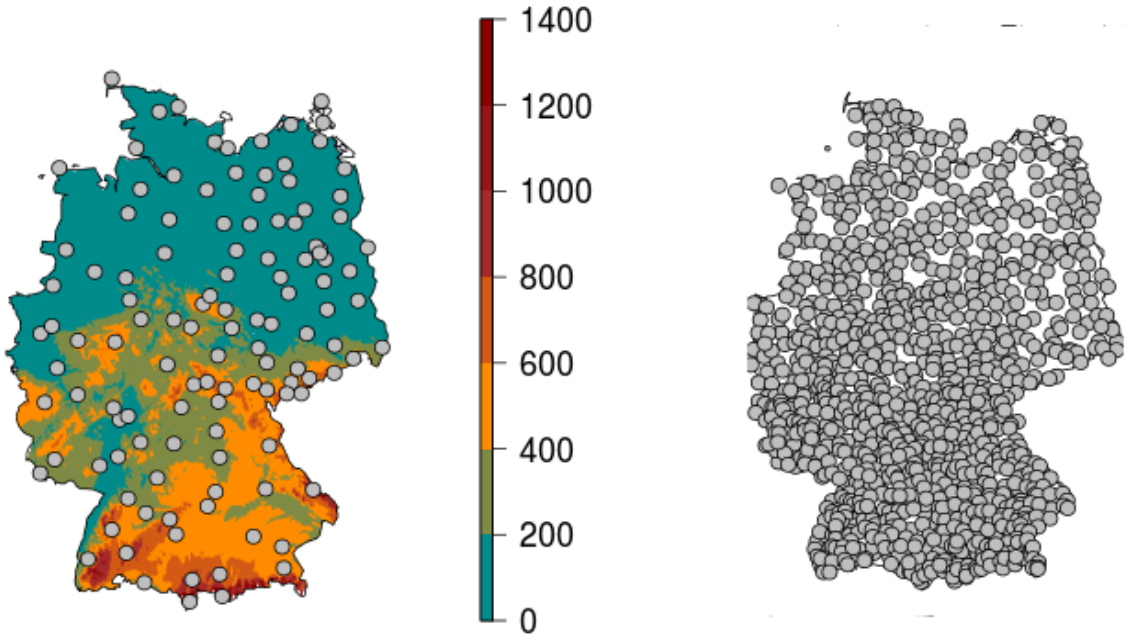

Figure 3.2.: Synoptic stations used for (left) training and (right) validating the model. The colors in the left figure show the elevation in meters.

starts to melt from the outside into the inside. To the radar this particle will be interpreted as a big water droplet that is slowly falling, which is another source of uncertainty (Rinehart 1997). Although radars have the advantage to monitor larger areas in comparison with synoptic stations and present also higher temporal resolution, which enables them to detect storms that may be undetected by rain gauges, the large uncertainties in the rainfall estimates need to be considered when using such data.

3.0.6. The datasets used in the study

Hourly precipitation data from the observational network of DWD with measurements starting at least in 1950 is used to train the method. This amounts 121 stations (Figure 3.2 (left)). Stations with long term measurements are selected here, since the purpose of the project is to simulate precipitation for longer periods back in time. Another set of 1384 stations is used for validation (Figure 3.2

(right)). This dataset does not include the training stations. All the data undergo a basic quality control at DWD.

Infrared brightness temperature (IRBT) from the 10.8μm channel of Meteosat-10 is used as covariate. The data has approximately 4 km x 4 km resolution and is available every 15 minutes. Figure 3.3 summarizes how brigthness temperature is obtained from the amount of radiative energy that reaches the satellite.

In order to have IRBT and precipitation data in the same resolution, the minimum 15-minute IRBT inside every hour is used to represent the infrared information within that hour. This approach is based on the following. A precipitation event, in particular convective precipitation is most likely to occurr in the presence of high and thick clouds (Arkin *et al.* 1994). The higher the cloud, the lower the temperature of its top and hence, the IRBT. Thus, the precipitation occurrence inside an hour may be most associated with the minimum value of IRBT in that hour.

Although during winter low IRBT may be frequently associated with the non-precipitating cirrus clouds and not with the convective systems (Bellerby, Sun 2005), the performance of the method improves by using this information during the winter as well.

Lastly, a dataset of radar-derived rainfall composites are used for validation. These composites are obtained from a network of 16 Doppler weather radars covering Germany and operated by DWD. The radar scans are performed at every 5 minutes with 1 km spatial resolution and 1 degree resolution in the azimuth. The reflectivities are quality controled at every scan to avoid that false echoes are considered as precipitation. The measurements are interpolated on a grid with 2.8 km spatial resolution in order to produce the composites. Radar pixels at every scan with false echoes detected are flagged. These are then considered in the interpolation method as well. After the radar reflectivity is interpolated these values are converted into rainfall rates through 5 different $Z-R$ relationships based on the meteorological conditions. The radar composites are compared to satellite data usually over longer periods (as a month) to verify whether there are inconsistencies. Besides, composites grid points with significant precipitation amount where satellite data shows no clouds are masked (Stephan *et al.* 2008).

For comparison with the fields simulated with HIRAIN the composites are accumulated at every hour and aggregated into a 4 km resolution grid.

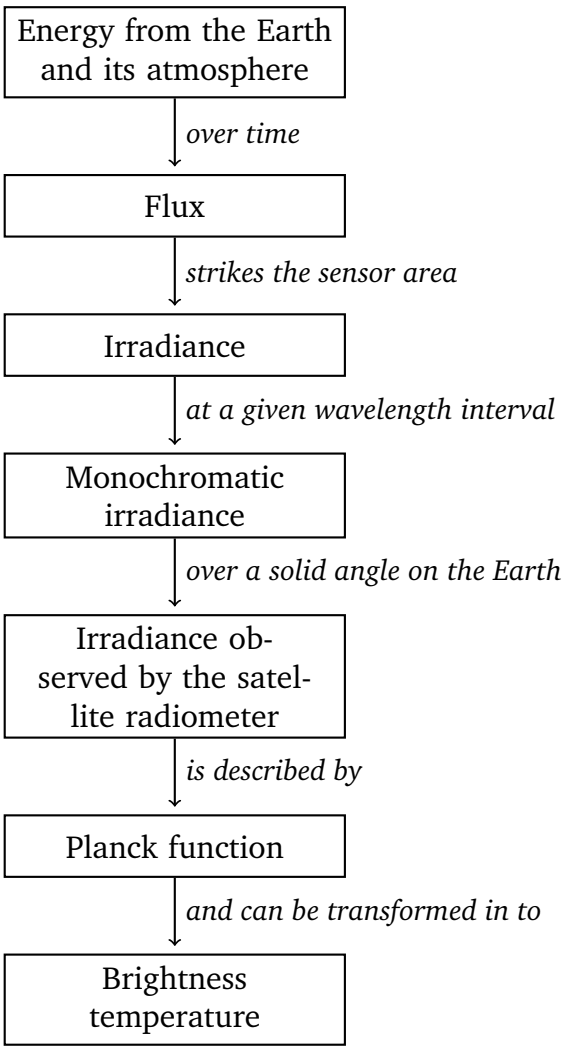

Figure 3.3.: Relationship between the energy from the Earth and its atmosphere that reaches the satellite sensors and the brightness temperature (Menzel Accessed 07 March 2016).

4. Bayesian spatial modeling

Basic concepts of spatial statistical modeling are reviewed in this chapter. The objective is to provide some background on the topic for a better understanding of the method used in this doctoral thesis, which is presented in chapter 5.

One common interest in the field of spatial data analysis is to predict a certain variable of interest in a space $A \subset \Re^2$ given the observations of that variable collected at a finite number of sites. It is then assumed that the observations are one possible realisation of an unknown underlying stochastic process, which is defined overall in that space. Although the observations are available only at a finite number of locations, they contain valuable information about the process that generated them. Based on that, spatial statistical models are used to infere the unknown spatial process from the datasets. The variable of interest can then be predicted at unobserved locations by drawing realisations from this process conditional on the observations.

One kind of statistical model commonly used to address this problem is a hierarchical model. This considers the existence of conditional dependences among the spatial process, the observations and the parameters. The latter may be parameters that describe the spatial structure of the process or that indicate the relationship between observations and covariates, when these are used. Sub-models given by conditional probability distributions can be specified for each of these quantities and then, treated hierarchically. If sub-models are specified for all the three quantities, the model is a Bayesian hierarchical model (BHM).

In order to understand the mechanism of a BHM, firstly an introduction on spatial stochastic process is given in section 4.1. Then, section 4.2 provides a more detailed overview of hierarchical models with a Bayesian approach.

4.1. Spatial stochastic process

4.1.1. Overview

A spatial stochastic process or also called spatial random field in some space $A \subset \Re^2$ is a collection of random variables S at locations $\mathbf{r} \in A$ and written as

$$\{S(\mathbf{r}) : \mathbf{r} \in A \subset \Re^2\}. \tag{4.7}$$

The simplest spatial stochastic process is the Gaussian process. This assumes that each random variable in the space $A \subset \Re^2$ is normally distributed. Such a process is completely described by a mean function, $\mu(\mathbf{r}) = E[S(\mathbf{r})]$ and a covariance function, $\gamma(\mathbf{r},\mathbf{r}') = Cov[S(\mathbf{r}),S(\mathbf{r}')]$ (Diggle, Ribeiro Jr. 2007). The mean function describes the expected value of the process and the covariance function describes how the values deviate from the mean.

A stochastic process $S(\mathbf{r})$ can be decomposed into three terms, namely a fixed effect $\mu(\mathbf{r})$, a random effect $U(\mathbf{r})$ and a random error $\varepsilon(\mathbf{r})$,

$$S(\mathbf{r}) = \mu(\mathbf{r}) + U(\mathbf{r}) + \varepsilon(\mathbf{r}). \tag{4.8}$$

The fixed effect $\mu(\mathbf{r})$ is given by the expected value and describes the deterministic nature of the process. It can be constant, $\mu(\mathbf{r}) = \mu$, variable in space $\mu(\mathbf{r}) = \mu(\mathbf{r})$ or depend on covariates, $\mu(\mathbf{r}) = \mathbf{d}^T(\mathbf{r})\boldsymbol{\beta}$, where $\mathbf{d}$ is a vector with the covariates at the locations $\mathbf{r}$ and $\boldsymbol{\beta}$ is the vector with the regression coefficients.

The remaining terms, $U(\mathbf{r})$ and $\varepsilon(\mathbf{r})$ describe the stochastic nature of the spatial process. They describe all the variability in the process that is not captured by the deterministic term. The random effect $U(\mathbf{r})$ describes the spatial dependency among the random variables in the space $A \subset \Re^2$ and the random error $\varepsilon(\mathbf{r})$ represents the errors due to microscale variations and measurement errors.

Commonly, in order to represent a spatial stochastic process with a statistical model, a single one is used on an entire domain. This implies that the process is assumed to be stationary. A spatial Gaussian process is stationary when the expected value $\mu(\mathbf{r})$ over the space $A \subset \Re^2$ is constant and the covariance depends only on the distance $\mathbf{h}$ between two locations $\mathbf{r}$ and $\mathbf{r}'$, i.e, $\mathbf{h}$ is the vector difference $\boldsymbol{r} - \boldsymbol{r}'$. Additionaly if the covariance depends only on the absolute value of the

distance between two locations and not on the direction, the process is stationary and isotropic (Diggle, Ribeiro Jr. 2007). In practice, however, these assumptions are usually relaxed and the process is considered intrinsic stationary when it is stationary in the covariance structure.

In order for the statistical model to represent a legitimate surface of the stationary spatial process, the following conditions need to be satisfied: a) the correlation in space between two locations needs to decrease as the distance between them increases, b) the correlation function needs to be continuous and smooth at every location in the domain $A \subset \Re^2$ and c) lastly, the correlation function needs to be positive definite (Diggle, Ribeiro Jr. 2007).

Spatial continuity assures that $S(\mathbf{r})$ is similar to the surroundings, otherwise, $S(\mathbf{r})$ can take any value. Mathematically this is described by the continuity and differentiability of the stochastic process. There are two kinds of continuity and differentiability, namely the mean square and the path continuity and differentiability. Diggle, Ribeiro Jr. 2007 discuss that for stationary Gaussian processthe mean square differentiability is an appropriate measure of the smothness.

A process is said to be mean square continuous if

$$E[\{S(\mathbf{r}+\mathbf{h}) - S(\mathbf{r})\}^2] \to 0 \tag{4.9}$$

as $\mathbf{h} \to 0$, and it is mean square differentiable if

$$E[\{\frac{S(\mathbf{r}+\mathbf{h}) - S(\mathbf{r})}{\mathbf{h}} - S'(\mathbf{r})\}^2] \to 0 \tag{4.10}$$

as $\mathbf{h} \to 0$.

Standard families of correlation functions that satisfy conditions a), b) and c) are available. This study uses the exponential function, which is a special case of the Matérn function.

The Matérn function reads,

$$\rho(\mathbf{h}) = (2^{\kappa-1}\Gamma(\kappa))^{-1}(\mathbf{h}/\phi)^{\kappa}K_{\kappa}(\mathbf{h}/\phi), \tag{4.11}$$

where $K_{\kappa}()$ is the modified Bessel function of order κ, and $\kappa > 0$ is a shape parameter, which determines the smoothness of the spatial process. The parameter $\phi > 0$ is a scale parameter that has the dimension of a distance. The spatial Gaus-

sian process defined by a Matérn covariance function is $\kappa_0 - 1$ times mean-square differentiable, where κ_0 is the smallest integer greater than or equal to κ (Diggle, Ribeiro Jr. 2007).

For $\kappa = 0.5$ the Matérn correlation function reduces to the exponential function,

$$\rho(\mathbf{h}) = exp(-\mathbf{h}/\phi). \tag{4.12}$$

4.1.2. Notation

In practice, a spatial stochastic process $S(\mathbf{r})$ is studied in terms of a finite number of random variables S at a finite number of locations $\mathbf{r}_i \in A$, where $i : 1,2,..,n$. In a statistical model this is represented by the random vector

$$\mathbf{S} = \begin{bmatrix} S(\mathbf{r}_1) \\ S(\mathbf{r}_2) \\ . \\ . \\ S(\mathbf{r}_n) \end{bmatrix} = \begin{bmatrix} S_1 \\ S_2 \\ . \\ . \\ S_n \end{bmatrix}, \tag{4.13}$$

and each possible realisation $\mathbf{s}$ at the same locations reads

$$\mathbf{s} = \begin{bmatrix} s_1 \\ s_2 \\ . \\ . \\ s_n \end{bmatrix}. \tag{4.14}$$

For a spatial process that is Gaussian, the joint distribution of the random variables $S(\mathbf{r}_1)$, $S(\mathbf{r}_2)$, ..., $S(\mathbf{r}_n)$ is multivariate normal. Thus, this spatial process follows a multivariate normal distribution, namely $\mathbf{S} \sim MVN(\boldsymbol{\mu}, \Sigma)$, where the vector

$\boldsymbol{\mu}$ contains the expected values at the locations $\mathbf{r}_i$,

$$E[\mathbf{S}] = \begin{bmatrix} E[S(\mathbf{r}_1)] \\ E[S(\mathbf{r}_2)] \\ . \\ . \\ E[S(\mathbf{r}_n)] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ . \\ . \\ \mu_n \end{bmatrix} = \boldsymbol{\mu}. \tag{4.15}$$

and Σ is the covariance matrix, which provides information of the variance due to the spatial process itself and due to microscale variations and measurement errors. Σ is given as

$$\sigma^2 R(\phi) + \tau^2 I. \tag{4.16}$$

The parameter σ^2 is the variance of the spatial process and it is constant when the stationarity is the case. The term $R(\phi)$ is a matrix with the spatial correlations of the random variables. The correlation depends on the parameter ϕ, which in turn indicates the distance at which two random variables are no longer correlated. This matrix reads,

$$R(\phi) = \begin{vmatrix} \rho(\boldsymbol{r}_1, \boldsymbol{r}_1; \phi) & \rho(\boldsymbol{r}_1, \boldsymbol{r}_2; \phi) & \dots & \rho(\boldsymbol{r}_1, \boldsymbol{r}_n; \phi) \\ \rho(\boldsymbol{r}_2, \boldsymbol{r}_1; \phi) & \rho(\boldsymbol{r}_2, \boldsymbol{r}_2; \phi) & \dots & \rho(\boldsymbol{r}_2, \boldsymbol{r}_n; \phi) \\ \dots & \dots & \dots & \dots \\ \rho(\boldsymbol{r}_n, \boldsymbol{r}_1; \phi) & \rho(\boldsymbol{r}_n, \boldsymbol{r}_2; \phi) & \dots & \rho(\boldsymbol{r}_n, \boldsymbol{r}_n; \phi) \end{vmatrix}, \tag{4.17}$$

where the entries $\rho(\boldsymbol{r}_i, \boldsymbol{r}_j; \phi)$ are given by a correlation function.

The term τ^2 is called nugget and provides the variance due to measurement errors and microscale variations. It is commonly assumed that these errors are not correlated in space. Hence, in this case the correlation matrix is the identity matrix I.

4.2. Bayesian hierarchical modeling

Having reviewed the basic concepts of spatial stochastic processes and the features of the Gaussian process, an introduction to hierarchical models is provided here.

Firstly, some concepts are presented of the probability theory that form the core

of such models. This is followed by an overview of hierarchical models in general and the section closes with the focus on the Bayesian approach. For simplicity the terms probability distribution and distribution will be used interchangeably from this point forward.

4.2.1. Basic concepts from the probability theory

A probability distribution of a random variable will be represented here by square brackets $[.]$. If A is a random variable, then the probability distribution of A is $[A]$.

Considering for instance two random variables, A and B, the conditional probability of A given B is represented as $[A|B]$ and writes

$$[A|B] = \frac{[B|A][A]}{\int [B|A][A]dA} = \frac{[B|A][A]}{[B]}, \tag{4.18}$$

where $[A]$ is also called the marginal distribution.

This is the Bayes' Theorem and basically forms the core of hierarchical models.

The joint probability distribution of A and B, namely $[A,B]$ can be written in terms of a conditional and a marginal distribution as

$$[A,B] = [A|B][B] = [B|A][A]. \tag{4.19}$$

By following the same logic, the joint probability distribution of n_r randon variables can be factored as a product of $n_r - 1$ conditional distributions and one marginal distribution. Considering for instance a third random variable C, the joint probability distribution of A, B and C is given as

$$[A,B,C] = [A|B,C][B|C][C]. \tag{4.20}$$

From the joint probability distribution, the marginal distribution of a random variable can be obtained by marginalizing the joint distribution over all other random variables. Given the joint distribution of A and B, the marginal distribution of A is given as

$$[A] = \int [A,B]dB = \int [A|B][B]dB. \tag{4.21}$$

Considering the joint distribution of A, B and C, the marginal distribution of A is

then given as

$$[A] = \int\int [A,B,C]dBdC = \int\int [A|B,C][B|C][C]dBdC. \tag{4.22}$$

4.2.2. Hierarchical modeling

In spatial statistical modeling there are at least three quantities of interest, namely the unknown spatial stochastic process S, the observations of a random variable Y and the unknown model's parameters θ.

Probability distributions are attributed to those quantities and then treated hierarchically. Based on which quantities are modeled by probability distributions, the models are classified either as empirical or as Bayesian (Cressie, Wikle 2011). A brief overview of both approaches is presented below.

In an empirical hierarchical model (EHM) one sub-model is specified for the process S and another for the random variable Y,

$$[\boldsymbol{y}^o, \boldsymbol{s}^o|\boldsymbol{\theta}] = [\boldsymbol{y}^o|\boldsymbol{s}^o, \boldsymbol{\theta}][\boldsymbol{s}^o|\boldsymbol{\theta}], \tag{4.23}$$

where the conditional probability distributions on the right hand side are the

$$\begin{cases} \text{Data model:} & [\boldsymbol{y}^o|\boldsymbol{s}^o, \boldsymbol{\theta}] \\ \text{Process model:} & [\boldsymbol{s}^o|\boldsymbol{\theta}]. \end{cases} \tag{4.24}$$

The vectors presented in the model are as follow. The observations of the variable of interest Y at the locations $\boldsymbol{r}_i^o, i = 1,2,\ldots,n$ in a space $A \subset \Re^2$ are represented by the vector

$$\boldsymbol{y}^o = (y(\boldsymbol{r}_1^o), y(\boldsymbol{r}_2^o), \ldots, y(\boldsymbol{r}_n^o))^T = (y_1^o, y_2^o, \ldots, y_n^o)^T. \tag{4.25}$$

The stochastic process S at the same locations is given by the vector

$$\boldsymbol{s}^o = (s(\boldsymbol{r}_1^o), s(\boldsymbol{r}_2^o), \ldots, s(\boldsymbol{r}_n^o))^T = (s_1^o, s_2^o, \ldots, s_n^o)^T, \tag{4.26}$$

and lastly, the vector $\boldsymbol{\theta} = [\theta_1, \theta_2, ..., \theta_k]^T$, in turn, contains k model's parameters.

In this approach uncertainties of the parameters are not considered. The parameters are estimated by some estimation algorithm, as the maximum likelihood

and the estimation is plugged in the process and in the data sub-models (4.24) as a fixed quantity. Thus, $\boldsymbol{\theta}$ is replaced by its estimator $\hat{\boldsymbol{\theta}}$ in those probability distributions.

A Bayesian hierarchical model (BHM) is a three-level model. Beside the two sub-models considered in the empirical approach, a third sub-model for the parameters in $\boldsymbol{\theta}$ is specified. The model is then given as

$$[\boldsymbol{y}^o, \boldsymbol{s}^o, \boldsymbol{\theta}] = [\boldsymbol{y}^o|\boldsymbol{s}^o, \boldsymbol{\theta}][\boldsymbol{s}^o|\boldsymbol{\theta}][\boldsymbol{\theta}] \tag{4.27}$$

where the right hand side is as follows,

$$\begin{cases} \text{Data model:} & [\boldsymbol{y}^o|\boldsymbol{s}^o, \boldsymbol{\theta}] \\ \text{Process model:} & [\boldsymbol{s}^o|\boldsymbol{\theta}] \\ \text{Parameter model:} & [\boldsymbol{\theta}]. \end{cases} \tag{4.28}$$

The uncertainties in the parameters are considered in this approach. The parameter model is a probability distribution that needs to be specified by the user and should give information on $\boldsymbol{\theta}$ before considering the data collection. This distribution is then called the prior distribution.

This study used the Bayesian approach, in order to consider the uncertainties of the parameters in the model and in the predictions. The next subsection will threrefore concentrate only on this approach.

4.2.3. Bayesian approach

A BHM aims at estimating the joint probability distribution of all observed and unobserved quantities (4.27). For that, two steps are performed. First, the posterior distribution of $\boldsymbol{\theta}$ and S is inferred given the observations $\boldsymbol{y}^o$, namely the distribution $[\boldsymbol{s}^o, \boldsymbol{\theta}|\boldsymbol{y}^o]$. In the second step, given the posterior of these quantities, predictions of $\boldsymbol{s}^p$ (the spatial process S at the unobserved locations $\boldsymbol{r}_j^p, j = 1, 2, \ldots, m$) are provided. This is obtained by simulating from the posterior predictive distribution $[\boldsymbol{s}^p|\boldsymbol{y}^o]$.

The posterior distribution is derived by rewriting (4.27) and applying the Bayes'

Theorem,

$$[\boldsymbol{s}^o, \boldsymbol{\theta}|\boldsymbol{y}^o] = \frac{[\boldsymbol{y}^o|\boldsymbol{s}^o, \boldsymbol{\theta}][\boldsymbol{s}^o|\boldsymbol{\theta}][\boldsymbol{\theta}]}{\int [\boldsymbol{y}^o|\boldsymbol{s}^o, \boldsymbol{\theta}][\boldsymbol{s}^o|\boldsymbol{\theta}][\boldsymbol{\theta}] d\boldsymbol{s}^o d\boldsymbol{\theta}} = \frac{[\boldsymbol{y}^o|\boldsymbol{s}^o, \boldsymbol{\theta}][\boldsymbol{s}^o|\boldsymbol{\theta}][\boldsymbol{\theta}]}{[\boldsymbol{y}^o]}. \tag{4.29}$$

Omiting $[\boldsymbol{y}^o]$ yields the unnormalized posterior distribution, which has the form

$$[\boldsymbol{s}^o, \boldsymbol{\theta}|\boldsymbol{y}^o] \propto [\boldsymbol{y}^o|\boldsymbol{s}^o, \boldsymbol{\theta}][\boldsymbol{s}^o|\boldsymbol{\theta}][\boldsymbol{\theta}]. \tag{4.30}$$

A similar logic is followed to derive the posterior predictive distribution $[\boldsymbol{s}^p|\boldsymbol{y}^o]$. From the probability theory, the marginal distribution of $\boldsymbol{y}^o$ is obtained by marginalizing the joint distribution $[\boldsymbol{s}^o, \boldsymbol{\theta}, \boldsymbol{y}^o]$ over $\boldsymbol{s}^o$ and $\boldsymbol{\theta}$,

$$[\boldsymbol{y}^o] = \int [\boldsymbol{y}^o, \boldsymbol{s}^o, \boldsymbol{\theta}] d\boldsymbol{s}^o d\boldsymbol{\theta} = \int [\boldsymbol{y}^o|\boldsymbol{s}^o, \boldsymbol{\theta}][\boldsymbol{s}^o, \boldsymbol{\theta}] d\boldsymbol{s}^o d\boldsymbol{\theta}. \tag{4.31}$$

Replacing the predictive distribution $[\boldsymbol{s}^p|\boldsymbol{y}^o]$ in (4.31), this can be expressed as

$$[\boldsymbol{s}^p|\boldsymbol{y}^o] = \int \int [\boldsymbol{s}^p|\boldsymbol{s}^o, \boldsymbol{\theta}, \boldsymbol{y}^o][\boldsymbol{s}^o, \boldsymbol{\theta}|\boldsymbol{y}^o] d\boldsymbol{s}^o d\boldsymbol{\theta} = \int \int [\boldsymbol{s}^p|\boldsymbol{s}^o, \boldsymbol{\theta}][\boldsymbol{s}^o, \boldsymbol{\theta}|\boldsymbol{y}^o] d\boldsymbol{s}^o d\boldsymbol{\theta}. \tag{4.32}$$

Since $\boldsymbol{s}^p$ and $\boldsymbol{y}^o$ are conditionally independent given $\boldsymbol{s}^o$ and $\boldsymbol{\theta}$, the second integral follows directly from the first.

If the integral in (4.32) can be directly calculated, $\boldsymbol{s}^p$ can be predicted by drawing samples from the posterior predictive distribution $[\boldsymbol{s}^p|\boldsymbol{y}]$. However, this is often not the case and therefore, simulating directly from the distribution is not possible. In such scenarios, Markov Chain Monte Carlo (MCMC) simulations are a useful tool.

Next section presents the basic concepts of MCMC simulations and how they are used to approximate unknown probability distributions.

4.2.4. Markov chain Monte Carlo (MCMC) algorithms

A Markov chain is a sequence of random variables, for instance $X_1, X_2, ..., X_t, ...$, for which, the distribution of X_t at any given moment t, depends only on the previous random variable X_{t-1}. This is known as the Markov property (Gilks *et al.* 1996).

A Markov chain starts from a point X_0 and is constructed by drawing samples from a transition or also callled proposal distribution $[X_t|X_{t-1}]$ that depends only

on the previous state of the chain X_{t-1}. The MCMC simulations aim at producing a Markov chain that converges to a stationary probability distribution. When MCMC algorithms are used to approximate a posterior probability distribution, this is then the stationary distribution.

Many algorithms have been developed to build MCMC sequences. In this section only a few of them will be presented, namely the Metropolis-Hasting (Metropolis *et al.* 1953; Hastings 1970), the Gibbs sampler (German, German 1984), and the the Langevin-Hastings (Christensen *et al.* 2001), which is the algorithm used in this study.

Considering that the Metropolis-Hastings algorithm is used to approximate the posterior distribution $[\boldsymbol{\theta}|\boldsymbol{y}^o]$. The starting point $\boldsymbol{\theta}_t = \boldsymbol{\theta}_0$ is drawn from a starting distribution $[\boldsymbol{\theta}_0]$. A candidate $\boldsymbol{\theta}^*$ for the next state $\boldsymbol{\theta}_{t+1}$ is sampled from the proposal distribution $[\boldsymbol{\theta}^*|\boldsymbol{\theta}_t]$ and the acceptance probability $\alpha(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t+1})$ is computed as

$$\alpha(\boldsymbol{\theta}_t, \boldsymbol{\theta}^*) = \min\left(1, \frac{[\boldsymbol{\theta}^*|\boldsymbol{y}^o][\boldsymbol{\theta}_t|\boldsymbol{\theta}^*]}{[\boldsymbol{\theta}_t|\boldsymbol{y}^o][\boldsymbol{\theta}^*|\boldsymbol{\theta}_t]}\right), \tag{4.33}$$

where $[.|\boldsymbol{y}^o]$ is the posterior distribution.

If $\alpha(\boldsymbol{\theta}_t, \boldsymbol{\theta}^*) = 1$ the next state in the chain is $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}^*$, otherwise a random number u is sampled from a uniform distribution $U(0,1)$ and

$$\boldsymbol{\theta}_{t+1} = \begin{cases} \boldsymbol{\theta}^* & \text{if u} < \alpha \\ \boldsymbol{\theta}_t & \text{if u} > \alpha \end{cases} \tag{4.34}$$

When the proposal distribution is symmetric, $[\boldsymbol{\theta}_{t+1}|\boldsymbol{\theta}_t] = [\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}]$ the method is known as the Metropolis algorithm and the acceptance probability reduces to

$$\alpha(\boldsymbol{\theta}_t, \boldsymbol{\theta}^*) = \min\left(1, \frac{[\boldsymbol{\theta}^*|\boldsymbol{y}^o]}{[\boldsymbol{\theta}_t|\boldsymbol{y}^o]}\right). \tag{4.35}$$

A random walk Metropolis is a special case of the Metropolis algorithm, for which the proposal has the form $[\boldsymbol{\theta}^*|\boldsymbol{\theta}_t] = [|\boldsymbol{\theta}_t - \boldsymbol{\theta}^*|]$ (Gilks *et al.* 1996).

It is expected that at each step of the simulation the distribution of the current draw becomes closer to the stationary distribution. In general, for the Markov chain to reach convergence, the simulations need to be run long enough. The number of iterations needed is commonly problem dependent. There are many

diagnostic tests that can be employed to investigate the convergence of the chain. More on that topic is presented in the next section.

The fact that a draw in a MCMC sequence depends on the previous state makes the chain autocorrelated. In practice, to obtain a sequence of independent random variables, only every *dth* draw is retained. Besides that, the initial portion of the chain generated before the convergence is achieved is also discarded. This is called the burn-in.

Another approach to build MCMC sequences is the Gibbs sampler. The simulation at each step cycles through all the b parameter subvectors in $\boldsymbol{\theta} = (\theta^1, \theta^2, ..., \theta^b)$. The proposal distribution is the full conditional distribution of each parameter in $\boldsymbol{\theta}$. In this case, the acceptance probability is always 1 and therefore every draw is accepted. The full conditional distribution of a parameter θ^i contained in $\boldsymbol{\theta}$ is its distribution given all the other parameters and the observations, namely

$$[\theta^i_{t+1} | \boldsymbol{\theta}^{-i}_t, \boldsymbol{y}^o], \tag{4.36}$$

where

$$\boldsymbol{\theta}^{-i}_t = (\theta^1_{t+1}, ..., \theta^{i-1}_{t+1}, \theta^{i+1}_t, ..., \theta^b_t) \tag{4.37}$$

contains all the components in the vector $\boldsymbol{\theta}$ with their current values except for θ^i.

The method is carried out in the following steps. A starting point $\boldsymbol{\theta}_t = \boldsymbol{\theta}_0$ for all unknown parameters is drawn from a starting distribution $[\boldsymbol{\theta}_0]$. A sample θ^i_{t+1} from the full conditional distribution $[\theta^i_{t+1} | \boldsymbol{\theta}^{-i}_t, \boldsymbol{y}^o]$ is drawn. Every new value θ^i_{t+1} is updated in the full conditional distribution and used for the next draw. The method cycles through all the parameters until the chain of every element in $\boldsymbol{\theta}$ has converged (Gelman *et al.* 2004).

The third algorithm presented here is the Langevin-Hastings method. It uses as proposal a multivariate normal distribution with gradient information from the posterior distribution in the mean. This directs the algorithm towards regions of high posterior density and for spatial statistics this algorithm has shown better performance than other MCMC methods (Waagepetersen *et al.* 2008).

Considering that Langevin-Hastings is used to approximate the posterior dis-

tribuiton $[\boldsymbol{s}^o|\boldsymbol{y}^o]$, the proposal distribution has the form

$$MVN(\boldsymbol{s}^o + (\delta/2)\frac{d\log[\boldsymbol{s}^o|\boldsymbol{y}^o]}{d\boldsymbol{s}^o}, \delta I) \tag{4.38}$$

where $\frac{d\log[\boldsymbol{s}^o|\boldsymbol{y}^o]}{d\boldsymbol{s}^o}$ is the gradient of the log-posterior distribution. The parameter δ is a user-specified proposal variance.

Commonly in this approach the components of $\boldsymbol{s}^o$ are updated simultaneously. MCMC schemes that update the components of $\boldsymbol{s}^o$ in turn have shown poor mixing of the chains because the components are correlated (Waagepetersen *et al.* 2008). The acceptance probability in this algorithm is the same as in the Metropolis-Hastings (4.33).

4.2.5. MCMC convergence diagnostics

An important practical matter when using MCMC algorithms is to know when the chain has reached stationarity. Many convergence diagnostics have been proposed to investigate this. In this section an overview is provided of some diagnoctics that are used in this study, namely the trace plots, autocorrelation plots, Geweke's diagnostic and the Heidelberg and Welch's diagnostic.

The trace plots are simply a scatter plot of the samples versus the number of iterations. They provide visual information of the mixing and of the stationarity of the chain. A trace plot that shows jumps from one point of the posterior distribution to another in an opposite region in very few iterations indicates good mixing of the chain. Stationarity can be checked through the evolution of the mean and the variance. If they do not change much with the number of iterations it is usually a sign that the chain has reached stationarity. Although, such plots are an easy and fast way to obtain this information, they are usually insufficient to proof convergence and further tests are commonly required.

The autocorrelation plots show the dependence among draws. Poor mixing of the chain is indicated through high autocorrelations between draws with large lags.

The Geweke's diagnostics (Geweke 1992) is a statistical test of the difference of the means between the first and the last part of the chain. For the null hypothesis it is assumed that both samples are from the same distribution and thus both means

are equal, i.e., $H_0 : \mu_1 = \mu_2$. If this is true the test statistic follows a standard normal distribution. The test compares the first 10% with the last 50% of the chain assuming that both parts are independent. Z-scores and p-values are computed. A significance level α is chosen and if the p-value $< \alpha$, the null hypothesis is rejected. Rejection in this case means that both parts of the chain are not from the same distribution. Hence, the chain has not reached convergence.

The Heidelberg and Welch's diagnostic (Heidelberger, Welch 1983 and Heidelberger, Welch 1981) is divided in two parts. First, the stationarity is evaluated and second, a half-width test is applied to the part of the chain that passes the first test. In the first test the null hypothesis assumes that the samples come from a stationary distribution. It uses the Cramer-von-Mises statistic test that is first applied to the entire chain. This test is successively applied after removing 10%, 20%, 30%, ..., etc of the chain until the null hypothesis is accepted or 50% of the chain is removed. The latter indicates that the samples are not from a stationary distribution and the MCMC simulation needs to be run longer. If the test is passed, the number of samples to remove and the number of samples to retain are indicated. The second part investigates whether the sample size that passed the first test is large enough to accurately estimate the mean. To this end, the mean is computed with the chain portion that passed the first test and a second mean is computed using only half of this chain portion. If the ratio between both means is lower than a chosen target value, the test is passed. Otherwise the MCMC simulations need to be run longer.

5. HIRAIN: a method to generate high resolution precipitation product

This chapter presents the HIRAIN method to generate high resolution precipitation product.

HIRAIN is carried out in two steps. Firstly, a BHM simulates at unobserved locations the latent spatial Gaussian process assumed to drive the occurrence of precipitation exceeding a certain threshold. Second, realisations of occurrence/non-occurrence of precipitation above the same threshold is provided by two different methodologies, given the simulations from the model.

This chapter presents the general ideia behind HIRAIN and chapter 6 discusses its specific setup.

5.1. Assumptions

The definitions of the quantities involved in the method are presented before HIRAIN is described.

The observations and the stochastic process at the observation locations $\boldsymbol{r}_i^o$ are defined as in (4.25) and (4.26), respectively. The covariate at the same locations is represented by the vector

$$\boldsymbol{d}^o = (d(\boldsymbol{r}_1^o), d(\boldsymbol{r}_2^o), \ldots, d(\boldsymbol{r}_n^o))^T = (d_1^o, d_2^o, \ldots, d_n^o)^T. \tag{5.39}$$

At the unobserved locations $\boldsymbol{r}_j^p$ the spatial process is given as

$$\boldsymbol{s}^p = (s(\boldsymbol{r}_1^p), s(\boldsymbol{r}_2^p), \ldots, s(\boldsymbol{r}_m^p))^T = (s_1^p, s_2^p, \ldots, s_m^p)^T, \tag{5.40}$$

the covariate as

$$\boldsymbol{d}^p = (d(\boldsymbol{r}_1^p), d(\boldsymbol{r}_2^p), \ldots, d(\boldsymbol{r}_m^p))^T = (d_1^p, d_2^p, \ldots, d_m^p)^T \tag{5.41}$$

and the variable to be predicted as

$$\boldsymbol{y}^p = (y(\boldsymbol{r}_1^p), y(\boldsymbol{r}_2^p), \ldots, y(\boldsymbol{r}_m^p))^T = (y_1^p, y_2^p, \ldots, y_m^p)^T, \tag{5.42}$$

which is the occurrence/non-occurrence of precipitation above a certain threshold. Lastly, $\boldsymbol{\theta}$ is the vector containing the model's parameters.

In the first step, HIRAIN simulates the latent spatial process at the unobserved locations, namely $\boldsymbol{s}^p$ with a Bayesian spatial generalized linear model (BSGLM) adapted from the package "geoRglm" (Christensen, Ribeiro Jr. 2002) written in the R statistical programming language (R Core Team 2013).

The BSGLM is based on the following considerations. The random responses y_i^o are assumed to follow an independent Bernoulli distribution with outcomes of either 0 or 1. The underlying latent spatial process that drives the observations is assumed to be Gaussian with isotropic and stationary covariance structure. The response variables relate to the spatial process through the logit function,

$$y_i^o \sim \text{Bernoulli}(\pi(\boldsymbol{r}_i^o) = \text{logit}^{-1}(s(\boldsymbol{r}_i^o))) \quad \text{for } i = 1, \ldots, n, \tag{5.43}$$

with the expected value of the responses equal to the probability $\pi(\boldsymbol{r}_i^o)$ of the outcome $y_i^o = 1$,

$$E[y_i^o | s(\boldsymbol{r}_i^o)] = \pi(\boldsymbol{r}_i^o) = \frac{\exp(s(\boldsymbol{r}_i^o))}{1 + \exp(s(\boldsymbol{r}_i^o))}. \tag{5.44}$$

It is assumed in the model that there is no nugget. The spatial stochastic process is then decomposed as

$$S(\boldsymbol{r}_i^o) = \mu(\boldsymbol{r}_i^o) + U(\boldsymbol{r}_i^o), \tag{5.45}$$

where the deterministic part $\mu(\boldsymbol{r}_i^o)$ is given by the linear predictor $\boldsymbol{d}^T(\boldsymbol{r}_i^o)\boldsymbol{\beta}$, and the random effect follows a multivariate normal distribution, notably $U(\boldsymbol{r}_i^o) \sim MVN(0, \Sigma)$. The elements of the covariance matrix Σ are given by $\sigma^2 \rho(\boldsymbol{r}_i, \boldsymbol{r}_j; \phi)$, where the correlation function is $\rho(\boldsymbol{r}_i, \boldsymbol{r}_j; \phi) = exp(-h/\phi)$ and h is the great circle distance between two locations

$$h = r_E \cos^{-1}(\sin\varphi_1 \sin\varphi_2 + \cos\varphi_1 \cos\varphi_2 \cos(\lambda_1 - \lambda_2)). \tag{5.46}$$

The parameter $r_E = 6.378$ km is the radius of the Earth considering the Earth to

be a sphere. The variables (λ_i, φ_i) define a location r_i in degree of longitude and latitude, respectively.

5.2. The method

The goal of HIRAIN is to predict occurrence/non-occurrence of precipitation above a threshold at unobserved locations, $\boldsymbol{y}^p$, given the observations $\boldsymbol{y}^o$ and covariates at both observed and unobserved locations, $\boldsymbol{d}^o$ and $\boldsymbol{d}^p$. HIRAIN thus simulates from the predictive distribution $[\boldsymbol{y}^p|\boldsymbol{y}^o,\boldsymbol{d}^o,\boldsymbol{d}^p]$.

Further quantities involved in HIRAIN are the model's parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \phi)$ and the spatial latent process at the observation and the prediction locations, i.e., $\boldsymbol{s}^o$ and $\boldsymbol{s}^p$, respectively. All these quantities follow the conditional independence structure presented in Figure 5.1. Quantities that are not connected by a line are conditionally independent given the random variables at all other nodes (Diggle, Ribeiro Jr. 2007).

The predictive distribution can be written as

$$
\begin{aligned}
[\boldsymbol{y}^p|\boldsymbol{y}^o,\boldsymbol{d}^o,\boldsymbol{d}^p] &= \int\int\int [\boldsymbol{y}^p|\boldsymbol{y}^o,\boldsymbol{s}^o,\boldsymbol{\theta},\boldsymbol{d}^o,\boldsymbol{s}^p,\boldsymbol{d}^p][\boldsymbol{y}^o,\boldsymbol{s}^o,\boldsymbol{\theta},\boldsymbol{d}^o,\boldsymbol{s}^p,\boldsymbol{d}^p|\boldsymbol{y}^p]d\boldsymbol{\theta}d\boldsymbol{s}^o d\boldsymbol{s}^p \\
&= \int\int\int [\boldsymbol{y}^p|\boldsymbol{s}^p][\boldsymbol{s}^p|\boldsymbol{y}^o,\boldsymbol{s}^o,\boldsymbol{\theta},\boldsymbol{d}^o,\boldsymbol{d}^p,\boldsymbol{y}^p][\boldsymbol{y}^o,\boldsymbol{s}^o,\boldsymbol{\theta},\boldsymbol{d}^o,\boldsymbol{d}^p]d\boldsymbol{\theta}d\boldsymbol{s}^o d\boldsymbol{s}^p \\
&= \int\int\int [\boldsymbol{y}^p|\boldsymbol{s}^p][\boldsymbol{s}^p|\boldsymbol{s}^o,\boldsymbol{\theta},\boldsymbol{d}^p,\boldsymbol{y}^p][\boldsymbol{s}^o|\boldsymbol{y}^o,\boldsymbol{\theta},\boldsymbol{d}^o,\boldsymbol{d}^p][\boldsymbol{y}^o,\boldsymbol{\theta},\boldsymbol{d}^o,\boldsymbol{d}^p]d\boldsymbol{\theta}d\boldsymbol{s}^o d\boldsymbol{s}^p \\
&= \int\int\int [\boldsymbol{y}^p|\boldsymbol{s}^p][\boldsymbol{s}^p|\boldsymbol{s}^o,\boldsymbol{\theta},\boldsymbol{d}^p][\boldsymbol{s}^o|\boldsymbol{y}^o,\boldsymbol{\theta},\boldsymbol{d}^o][\boldsymbol{\theta}|\boldsymbol{y}^o,\boldsymbol{d}^o,\boldsymbol{d}^p]d\boldsymbol{\theta}d\boldsymbol{s}^o d\boldsymbol{s}^p \\
&= \int\int\int [\boldsymbol{y}^p|\boldsymbol{s}^p][\boldsymbol{s}^p|\boldsymbol{s}^o,\boldsymbol{\theta},\boldsymbol{d}^p][\boldsymbol{s}^o|\boldsymbol{y}^o,\boldsymbol{\theta},\boldsymbol{d}^o][\boldsymbol{\theta}]d\boldsymbol{\theta}d\boldsymbol{s}^o d\boldsymbol{s}^p, \qquad (5.47)
\end{aligned}
$$

where some conditional distributions are simplified based on the conditional independence structure in Figure 5.1.

The vector $\boldsymbol{\theta}$ contains the model's parameters $\boldsymbol{\beta}$, σ^2 and ϕ. $[\boldsymbol{\theta}]$ is the prior probability distribution. By using a conjugate prior for the parameters $\boldsymbol{\beta}$ and σ^2, the influence of $\boldsymbol{\beta}$ and σ^2 can be integrated out analytically by marginalizing over $\boldsymbol{\beta}$ and σ^2. The marginalization over $\boldsymbol{\beta}$ and σ^2 leads to which is a very complex integral and can not be solved analytically.

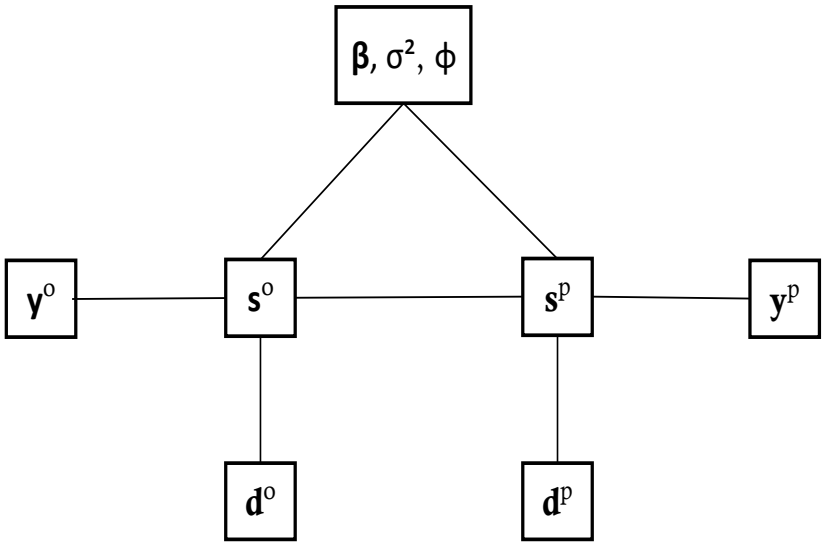

Figure 5.1.: Conditional independence structure of HIRAIN.

$$[\boldsymbol{y}^p|\boldsymbol{y}^o,\boldsymbol{d}^o,\boldsymbol{d}^p] = \int\int\int [\boldsymbol{y}^p|\boldsymbol{s}^p][\boldsymbol{s}^p|\boldsymbol{s}^o,\phi,\boldsymbol{d}^p][\boldsymbol{s}^o|\boldsymbol{y}^o,\phi,\boldsymbol{d}^o][\phi]d\phi d\boldsymbol{s}^o d\boldsymbol{s}^p$$
$$= \int\int\int [\boldsymbol{y}^p|\boldsymbol{s}^p][\boldsymbol{s}^p|\boldsymbol{s}^o,\phi,\boldsymbol{d}^p][\boldsymbol{s}^o,\phi|\boldsymbol{y}^o,\boldsymbol{d}^o]d\phi d\boldsymbol{s}^o d\boldsymbol{s}^p, \qquad (5.48)$$

Thus, simulating directly from $[\boldsymbol{y}^p|\boldsymbol{y}^o,\boldsymbol{d}^o,\boldsymbol{d}^p]$ is not possible. In order to obtain realisations of $\boldsymbol{y}^p$ HIRAIN follows in two steps and adopts the following procedures.

In step 1 of HIRAIN the BSGLM estimates $[\boldsymbol{s}^p,\phi|\boldsymbol{y}^o,\boldsymbol{d}^o]$ and draws samples of $\mathbf{s}^o$ and ϕ. The posterior predictive distribution $[\boldsymbol{s}^p|\boldsymbol{s}^o,\phi,\boldsymbol{d}^p]$ is then used to simulate $\boldsymbol{s}^p$ at unobserved locations. Thus, the final product of the BSGLM are the samples of $\boldsymbol{s}^p$ from (Diggle, Ribeiro Jr. 2007)

$$[\boldsymbol{s}^p|\boldsymbol{y}^o,\boldsymbol{d}^o,\boldsymbol{d}^p] = \int\int [\boldsymbol{s}^p|\boldsymbol{s}^o,\phi,\boldsymbol{d}^p][\boldsymbol{s}^o,\phi|\boldsymbol{y}^o,\boldsymbol{d}^o]d\phi d\boldsymbol{s}^o. \qquad (5.49)$$

In step 2, $\boldsymbol{y}^p$ is predicted conditionally on the simulations of $\boldsymbol{s}^p$. This part is per-

formed outside the BSGLM and employes two methodologies. One method draws samples from the first distribution in the integral 5.48, namely the Bernoulli distribution $[\boldsymbol{y}^p|\boldsymbol{s}^p]$. In a second approach values of $\boldsymbol{y}^p$ are generated by thresholding the simulated values of $\boldsymbol{s}^p$, assuming a deterministic relation between Y and S.

In the following, sections 5.3 and 5.4 describe in detail how step 1 and step 2 of HIRAIN are constructed.

5.3. Step 1: Predicting the latent process $\boldsymbol{s}^p$

5.3.1. Posterior distributions

5.3.1.1. Formulation

The posterior distribution of $\boldsymbol{s}^o$ and ϕ can be written as (compare conditional independence structure in Figure 5.1)

$$\begin{aligned}[\boldsymbol{s}^o,\phi|\boldsymbol{y}^o,\boldsymbol{d}^o] &\propto [\boldsymbol{y}^o|\boldsymbol{s}^o,\phi,\boldsymbol{d}^o][\boldsymbol{s}^o|\phi,\boldsymbol{d}^o][\phi] \\ &\propto [\boldsymbol{y}^o|\boldsymbol{s}^o][\boldsymbol{s}^o|\phi,\boldsymbol{d}^o][\phi],\end{aligned} \tag{5.50}$$

where the likelihood reads $[\boldsymbol{y}^o|\boldsymbol{s}^o] = \prod_{i=1}^n [y_i^o|\pi(\boldsymbol{r}_i^o)] = \prod_{i=1}^n \pi(\boldsymbol{r}_i^o)^{y_i^o}(1-\pi(\boldsymbol{r}_i^o))^{1-y_i^o}$ due to the assumption 5.43. For the prior distribution $[\phi]$ an inverse Gamma discrete distribution is specified,

$$[\phi] = \frac{\beta_g^{\alpha_g}\phi^{-(\alpha_g+1)}e^{-\beta_g/\phi}}{\Gamma(\alpha_g)}, \quad \phi > 0, \tag{5.51}$$

where α_g is the shape and β_g is the scale parameter.

The posterior distribution of the parameters $\boldsymbol{\beta}$ and σ^2 is known and has the form (Diggle, Ribeiro Jr. 2007)

$$\begin{aligned}[\boldsymbol{\beta},\sigma^2|\boldsymbol{s}^o,\phi] &\propto [\boldsymbol{s}^o|\boldsymbol{\beta},\sigma^2,\phi][\boldsymbol{\beta},\sigma^2|\phi] \\ &\quad [\boldsymbol{s}^o|\boldsymbol{\beta},\sigma^2,\phi][\boldsymbol{\beta}|\sigma^2,\phi][\sigma^2|\phi],\end{aligned} \tag{5.52}$$

where $[\boldsymbol{s}^o|\boldsymbol{\beta},\sigma^2,\phi]$ is the multivariate normal distribution

$$[\boldsymbol{s}^o|\boldsymbol{\beta},\sigma^2,\phi] \propto |\sigma^2 R_o(\phi)|^{-1/2} \exp\left(\frac{-1}{2}(\boldsymbol{s}^o - \boldsymbol{d}^o\boldsymbol{\beta})^T |\sigma^2 R_o(\phi)|^{-1}(\boldsymbol{s}^o - \boldsymbol{d}^o\boldsymbol{\beta})\right) \quad (5.53)$$

and $[\boldsymbol{\beta}|\sigma^2,\phi]$ and $[\sigma^2|\phi]$ are the following conjugate prior distributions

$$[\boldsymbol{\beta}|\sigma^2,\phi] \propto 1, \quad (5.54)$$

and a scaled inverse χ^2 distribution on σ^2,

$$[\sigma^2|\phi] = \frac{(S_\sigma^2 n_\sigma/2)^{n_\sigma/2}}{\Gamma(n_\sigma/2)} \sigma^{2(-n_\sigma/2+1)} \exp(-n_\sigma S_\sigma^2/2\sigma^2), \quad (5.55)$$

where n_σ and S_σ^2 are the degrees of freedom and scale parameters, respectively.

The substitution of these distributions on the right hand side of (5.52) leads to a normal scaled inverse χ^2 posterior distribution (Diggle, Ribeiro Jr. 2007) of the form

$$[\boldsymbol{\beta},\sigma^2|\boldsymbol{s}^o,\phi] \propto N - \chi^2_{sci}(\boldsymbol{\beta}', V_{\boldsymbol{\beta}'}, n_\sigma + n, S_1^2) \quad (5.56)$$

with

$$\boldsymbol{\beta}' = V_{\boldsymbol{\beta}'} \boldsymbol{d}^{oT} {R_o}^{-1} \boldsymbol{s}^o \quad (5.57)$$

$$V_{\boldsymbol{\beta}'} = (\boldsymbol{d}^{oT} {R_o}^{-1} \boldsymbol{d}^o)^{-1} \quad (5.58)$$

$$S_1^2 = \frac{n_\sigma S_\sigma^2 + \boldsymbol{s}^{oT} {R_o}^{-1} \boldsymbol{s}^o - \boldsymbol{\beta}'^T V_{\boldsymbol{\beta}'}^{-1} \boldsymbol{\beta}'}{n_\sigma + n}, \quad (5.59)$$

where n is the number of observation locations.

5.3.1.2. Simulating $\boldsymbol{s}^o$, $\boldsymbol{\beta}$, σ^2 and ϕ

Samples of $\boldsymbol{\beta}$ and σ^2 can directly be drawn from (5.56). Samples of $\boldsymbol{s}^o$ and ϕ, however, can not be directly drawn from their posterior. The likelihood in the right hand side of (5.50) consists of a high dimensional normalizing integral that can

not be solved analytically. Hence, a MCMC algorithm is employed to approximate $[\boldsymbol{s}^o, \phi|\boldsymbol{y}^o, \boldsymbol{d}^o]$.

Values of $\boldsymbol{s}^o$ and ϕ are simulated sequentially. At each new iteration $\boldsymbol{s}^o$ is updated given the current value of ϕ. In order to describe how the MCMC simulates $\boldsymbol{s}^o$ given the current ϕ, it will be considered for a moment that ϕ is fixed.

The algorithm used to update the spatial process is the Langevin-Hastings method and this approximates the distribution

$$[\boldsymbol{s}^o|\boldsymbol{y}^o] \propto [\boldsymbol{y}^o|\boldsymbol{s}^o][\boldsymbol{s}^o], \tag{5.60}$$

for a given ϕ. $[\boldsymbol{y}^o|\boldsymbol{s}^o]$ is the likelihood and $[\boldsymbol{s}^o]$ is obtained by marginalizing the joint distribution $[\boldsymbol{s}^o, \boldsymbol{\beta}, \sigma^2]$ over $\boldsymbol{\beta}$ and σ^2. $[\boldsymbol{s}^o]$ is an improper distribution given as (Christensen, Ribeiro Jr. 2002)

$$[\boldsymbol{s}^o] \propto |\boldsymbol{d}^{oT} R_o^{-1} \boldsymbol{d}^o|^{-1/2} |R_o|^{-1/2} n_\sigma S_\sigma \boldsymbol{s}^{oT} (R_o^{-1} - R_o^{-1} \boldsymbol{d}^o (\boldsymbol{d}^{oT} R_o^{-1} \boldsymbol{d}^o)^{-1} \boldsymbol{d}^{oT} R_o^{-1}) \boldsymbol{s}^{o\,-(n-p+n_\sigma)/2} \tag{5.61}$$

where $R_o = R_o(\phi)$ is the correlation matrix of $\boldsymbol{s}^o$ for a given ϕ.

Langevin-Hastings method may not work well if the components of $\boldsymbol{s}^o$ have very different posterior variances. The correlation structure of the proposal might be very different from the posterior correlation structure, where the components are not independent. Therefore, $\boldsymbol{s}^o$ is commonly reparameterised such that its components are less correlated a posteriori (Waagepetersen *et al.* 2008).

The reparameterisation in the BSGLM specifies that $\boldsymbol{s}^o = R_o^{0.5} \boldsymbol{\gamma}$, where $R_o^{0.5}$ is the Cholesky factorisation of the correlation matrix $R_o = R_o(\phi)$ and $\boldsymbol{\gamma} \sim N(0, I)$. Then, the Langevin-Hastings algorithm simulates from the posterior distribution $[\boldsymbol{\gamma}|\boldsymbol{y}^o]$ and samples from $[\boldsymbol{s}^o|\boldsymbol{y}^o]$ are obtained through the multiplication of the samples of $\boldsymbol{\gamma}$ by $R_o^{0.5}$ (Diggle, Ribeiro Jr. 2007). The proposal distribution used to update $\boldsymbol{\gamma}$ is a multivariate normal distribution, namely $MVN(\boldsymbol{\gamma} + (\delta/2)\nabla(\boldsymbol{\gamma})^{trunc}, \delta I)$, where δ is a user-specified proposal variance and should be tuned such that 60% acceptance rate is achieved (Waagepetersen *et al.* 2008). The term $\nabla(\boldsymbol{\gamma})^{trunc}$ is the truncated gradient of the log-posterior distribution of $\boldsymbol{\gamma}$. This is obtained as follows.

The posterior distribution of $\boldsymbol{\gamma}$ given the observations $\boldsymbol{y}^o$ is of the form

$$[\boldsymbol{\gamma}|\boldsymbol{y}^o] \propto \prod_{i=1}^{n} [y_i|R_o^{0.5}\gamma_i][\boldsymbol{\gamma}]. \tag{5.62}$$

Applying the logarithm to it, this reads (Christensen, Waagepetersen 2002)

$$\begin{aligned}\log[\boldsymbol{\gamma}|\boldsymbol{y}^o] \propto & const(y) + \sum_{i=1}^{n}\left(y_i^o R_o^{0.5}\gamma_i - \log(1+exp(R_o^{0.5}\gamma_i))\right) \\ & -\frac{1}{2}\boldsymbol{\gamma}^T(I_n - R_o^{-0.5}\boldsymbol{d}^o(\boldsymbol{d}^{oT}R_o^{-1}\boldsymbol{d}^o)^{-1}\boldsymbol{d}^{oT}R_o^{-0.5})\boldsymbol{\gamma}\end{aligned} \tag{5.63}$$

and the gradient of this logarithm gives

$$\begin{aligned}\nabla(\boldsymbol{\gamma})^{trunc} = \frac{\partial \log[\boldsymbol{\gamma}|\boldsymbol{y}^o]^{trunc}}{\partial \boldsymbol{\gamma}} = & R_o^{0.5^T}\left(y_i^o - \frac{\exp(s_i)}{1+\exp(s_i)}\right)_{i=1}^{n} \\ & -\left(I_n - R_o^{-0.5}\boldsymbol{d}^o(\boldsymbol{d}^{oT}R_o^{-1}\boldsymbol{d}^o)^{-1}\boldsymbol{d}^{oT}R_o^{-0.5^T}\right)\boldsymbol{\gamma},\end{aligned} \tag{5.64}$$

where the term $\left(y_i^o - \exp(s_i)/(1+\exp(s_i))\right)$ is bounded.

The acceptance probability reads

$$\alpha(\boldsymbol{\gamma}_t, \boldsymbol{\gamma}_{t+1}) = \frac{[\boldsymbol{\gamma}_{t+1}|\boldsymbol{y}^o]exp(-1/2\delta||\boldsymbol{\gamma}_t - m(\boldsymbol{\gamma}_{t+1})||^2)}{[\boldsymbol{\gamma}_t|\boldsymbol{y}^o]exp(-1/2\delta||\boldsymbol{\gamma}_{t+1} - m(\boldsymbol{\gamma}_t)||^2)}, \tag{5.65}$$

where $m(\boldsymbol{\gamma}) = \boldsymbol{\gamma} + (\delta/2)\nabla(\boldsymbol{\gamma})^{trunc}$ is the proposal's mean.

This procedure is repeated for every updated ϕ simulated through the random walk Metropolis algorithm. A Gaussian distribution with mean rounded to the nearest ϕ in the discrete prior set is used as proposal, namely $N(\phi, \delta_\phi I)$. The term δ_ϕ is the proposal variance, which must be provided by the user and should be tuned such that the acceptance rate is about 25% (Waagepetersen *et al.* 2008).

The sequential updating of $\boldsymbol{s}^o$ and ϕ produces samples from the posterior distribution $[\boldsymbol{s}^o, \phi|\boldsymbol{y}^o, \boldsymbol{d}^o]$.

5.3.2. Posterior predictive distribution

Going back to the predictive distribution in (5.48), the focus here will be on the posterior predictive distribution $[\boldsymbol{s}^p|\boldsymbol{s}^o,\phi,\boldsymbol{d}^p]$ given as

$$\begin{aligned}[\boldsymbol{s}^p|\boldsymbol{s}^o,\phi,\boldsymbol{d}^p] &= \int\int [\boldsymbol{s}^p|\boldsymbol{s}^o,\boldsymbol{\beta},\sigma^2,\phi,\boldsymbol{d}^p][\boldsymbol{\beta},\sigma^2|\boldsymbol{s}^o,\phi,\boldsymbol{d}^p]d\boldsymbol{\beta}d\sigma^2 \\ &= \int\int [\boldsymbol{s}^p|\boldsymbol{s}^o,\boldsymbol{\beta},\sigma^2,\phi,\boldsymbol{d}^p][\boldsymbol{\beta},\sigma^2|\boldsymbol{s}^o,\phi]d\boldsymbol{\beta}d\sigma^2. \end{aligned} \tag{5.66}$$

$[\boldsymbol{s}^p|\boldsymbol{s}^o,\boldsymbol{\beta},\sigma^2,\phi,\boldsymbol{d}^p]$ is a multivariate normal distribution with expected value

$$E[\boldsymbol{s}^p|\boldsymbol{s}^o,\boldsymbol{\beta},\sigma^2,\phi,\boldsymbol{d}^p] = \boldsymbol{d}^p\boldsymbol{\beta} + r^T R_o^{-1}(\boldsymbol{s}^o - \boldsymbol{d}^p\boldsymbol{\beta}) \tag{5.67}$$

and variance

$$Var[\boldsymbol{s}^p|\boldsymbol{s}^o,\boldsymbol{\beta},\phi,\sigma^2,\boldsymbol{d}^p] = \sigma^2(R_p - r^T R_o^{-1} r), \tag{5.68}$$

where $R_p = R_p(\phi)$ is the correlation matrix of $\boldsymbol{s}^p$, and $r = r(\phi)$ is the matrix with the cross-correlation between $\boldsymbol{s}^o$ and $\boldsymbol{s}^p$.

The distribution $[\boldsymbol{\beta},\sigma^2|\boldsymbol{s}^o,\phi]$ is the normal scaled inverse-χ^2 posterior in (5.56). The integration in (5.66) has an analytical solution and yields the following multivariate-t distribution (Diggle, Ribeiro Jr. 2007)

$$[\boldsymbol{s}^p|\boldsymbol{s}^o,\phi,\boldsymbol{d}^p] \sim t_{n+n_\sigma}(\boldsymbol{\mu}_1, S_1^2\Sigma_1), \tag{5.69}$$

where

$$\boldsymbol{\mu}_1 = r^T {R_o}^{-1}\boldsymbol{s}^o + (\boldsymbol{d}^p - r^T {R_o}^{-1}\boldsymbol{d}^o)\boldsymbol{\beta}', \tag{5.70}$$

$$\Sigma_1 = R_p - r^T {R_o}^{-1} r + (\boldsymbol{d}^p - r^T {R_o}^{-1}\boldsymbol{d}^o)V'_{\boldsymbol{\beta}}(\boldsymbol{d}^p - r^T {R_o}^{-1}\boldsymbol{d}^o)^T, \tag{5.71}$$

with $R_o = R_o(\phi)$ and S_1^2 is given by (5.59).

Simulating directly from this distribution provides samples of the stochastic process $\boldsymbol{s}^p$. The probability $\pi(\boldsymbol{r}^p)$ is obtained by applying (5.44) to these samples.

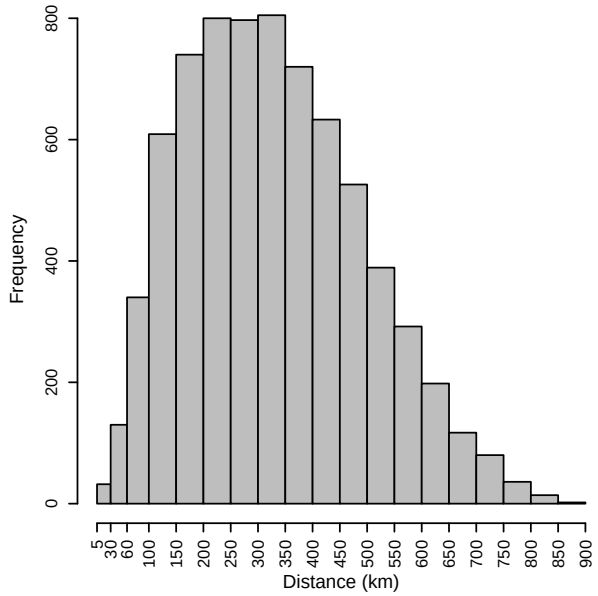

Figure 5.2.: Distance in km between all pairs of training synoptic stations.

5.4. Step 2: Generating precipitation occurrences y^p

Having simulated the latent Gaussian process with the BSGLM, two methodologies are investigated to generate occurrence/non-occurrence of precipitation above a threshold.

One approach is to simulate $\boldsymbol{y}^p$ as a spatially independent Bernoulli process $[\boldsymbol{y}^p|\boldsymbol{s}^p]$, which considers that the precipitation process is conditionally independent in space given $\boldsymbol{s}^p$. This is named here Bernoulli sampling (BNS) and is consistent with the assumption in step 1 that the observations $\boldsymbol{y}^o$ are generated by a spatially independent Bernoulli process. For neighbor distances between synoptic stations of about 5.6 km to 13.35 km, this is justified. However, since the predictions of $\boldsymbol{y}^p$ are carried out on a 4km grid, the dependence in precipitation between neighboring locations may not be ignored anymore.

An appropriate approach to generate $\boldsymbol{y}^p$ should account for spatial dependencies in the precipitation process. That is, after accounting for the spatial dependence in the latent Gaussian process, there is likely a 'residual' spatial dependence in the

occurrence/non-occurrence of precipitation.

The spatial dependence structure of the responses is unknown and would have to be inferred from the observations $\boldsymbol{y}^o$. However, there is relatively large distance between the observations in space. The minimum distance between two observation locations is 5.6 km. Figure 5.2 shows that most of the distances are greater than 60 km and the bulk of the distances is concentrated between 200 km and 350 km. This makes it challenging to infer a reliable dependence structure for the precipitation process. Further, MCMC simulations would be also required, making the estimation procedure even more time consuming. Including this dependence would greatly complicate inference and prohibit the use of 'geoRglm'.

For this reason, in a second approach $\boldsymbol{y}^p$ is generated by thresholding the realisations of $\boldsymbol{s}^p$. This is called the thresholding method (THR). All regions above an a priori defined threshold are considered to be regions with precipitation, whereas no precipitation is assigned elsewhere. THR implicitly assumes that the precipitation process is fully determined by the latent Gaussian process. This is inconsistent with the assumption in the BSGLM. However, the spatial dependence that is considered in the process might nevertheless provide more realistic fields of precipitation occurrence/non-occurrence.

6. HIRAIN Setup

This chapter discusses the setup of HIRAIN and how it is applied to many thresholds of precipitation amount to produce high resolution precipitation product.

The first part shows how the thresholds of precipitation are selected and the second part describes the structure of HIRAIN.

6.1. Selecting thresholds of precipitation

In order to select the thresholds that best represent the precipitation intensities in Germany, histograms are computed for hourly precipitation data from 1418 synoptic stations during the period from 2004 to 2011 (Figure 6.1).

The distributions show a common pattern during all seasons. Low intensity of precipitation occurs more frequently than high intensity. However, there are differences in the number of occurrences among the seasons.

The winter months (December-January-February (DJF)) show the highest frequencies of precipitation equal 0.1 mm/h and between 0.1 mm/h and 0.5 mm/h compared to the remaining seasons. Precipitation amounts $\geq$ 2.5 mm/h show the lowest frequencies of the year during DJF. This pattern is expected given that precipitation events are mostly originated from stratiform precipitating systems during this time.

During March-April-May (MAM) there is a change in the pattern when compared to DJF. The occurrence of precipitation equal 0.1 mm/h and between 0.1 mm/h and 0.5 mm/h starts decreasing, whereas for greater amounts ($\geq$ 2.5 m/h) it increases. Opposite to DJF, amounts between 5.0 mm/h and 7.0 mm/h are observed during MAM.

In June-July-August (JJA) the number of low intensity precipitation ($\leq$ 1.0 mm/h) decreases even more in comparison to DJF. Precipitation $\geq$ 2.5 mm/h

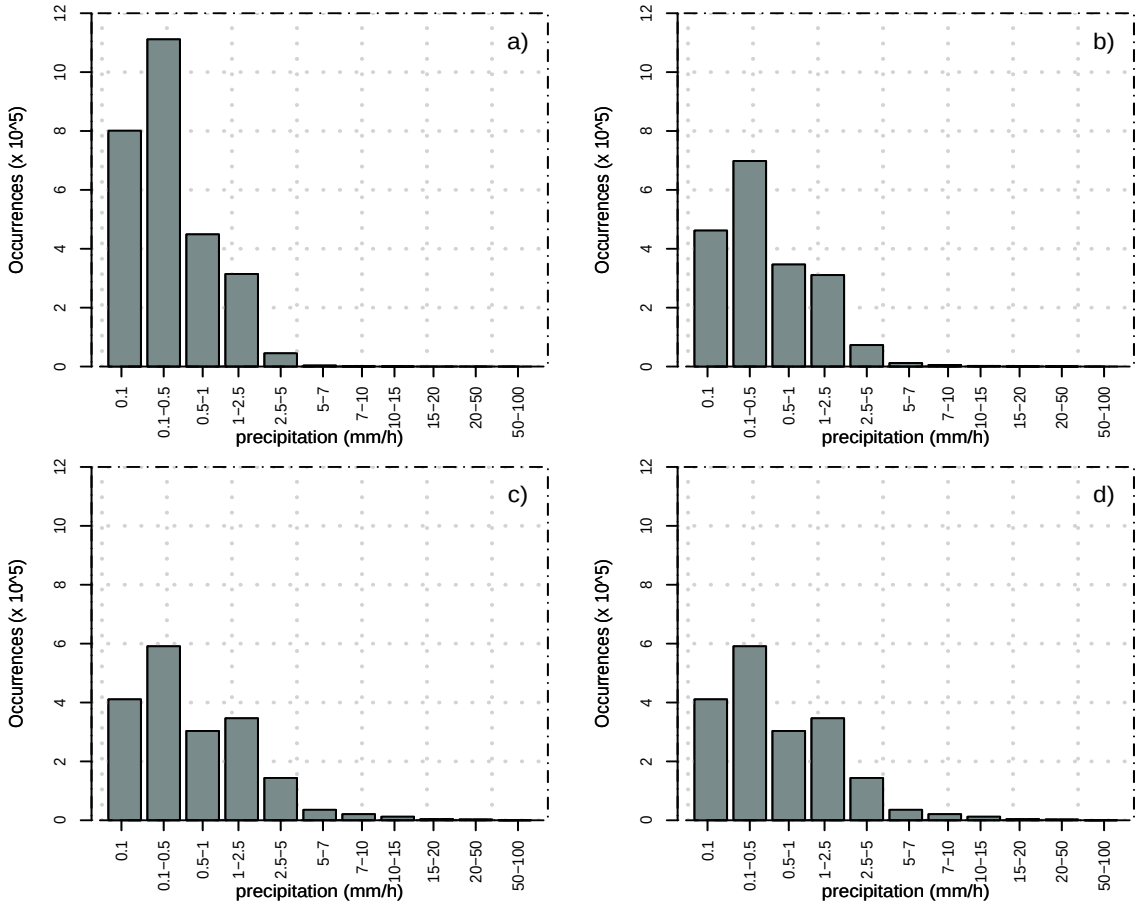

Figure 6.1.: Histogram of precipitation during (a) December-January-February (DJF), (b) March-April-May (MAM), (c) June-July-August (JJA) and (d) September-October-November (SON) based on hourly data from 1418 synoptic stations in the period from 2004 to 2011.

increases when compared to MAM and DJF. Amounts $\geq$ 15.0 mm/h are observed during this season. This is due to the more frequent convective precipitating systems, which might explain the increase in the amount when compared to DJF and MAM.

September-October-November (SON) show patterns that are very similar to JJA.

These observations suggest that precipitation in Germany ranges from intensities of 0.1 mm/h to about 7.0-10.0 mm/h for most events during the year. Extreme events are rarer and occur mostly during JJA and SON.

The classes of precipitation shown in the histograms seem to be a good representation of the precipitation intensities and their occurrences in Germany. The

thresholds of precipitation amount to be used in this study are chosen based on these classes and are the following: precipitation $\geq$ 0.1 mm/h, $\geq$ 0.5 mm/h, $\geq$ 1.0 mm/h, $\geq$ 2.5 mm/h, $\geq$ 5.0 mm/h, $\geq$ 7.0 mm/h and $\geq$ 10.0 mm/h.

6.2. Setup

HIRAIN is first applied to occurrence/non-occurrence of precipitation $\geq$ 0.1 mm/h. The structure of the method is schematically represented in the flow chart of Figure 6.2. HIRAIN runs at hourly time scale and for each time independently, i.e., the run for time t does not depend on time $t-1$.

The method starts by preparing the BSGLM's input data $\boldsymbol{y}^o$, $\boldsymbol{d}^o$, $\boldsymbol{d}^p$, $\boldsymbol{r}^o$, $\boldsymbol{r}^p$ for the first threshold $thr = 0.1$ mm/h and for the first time t_1. For convenience the time and threshold indices are omitted in the notation of the quantities in the flow chart.

The vector $\boldsymbol{y}^o$ contains 0's and 1's representing the occurrence/non-occurrence of precipitation $\geq$ 0.1 mm/h at the observation locations. The vectors $\boldsymbol{d}^o$ and $\boldsymbol{d}^p$ contain the covariate, namely IRBT at the observation and at the prediction locations, respectively. The corresponding coordinates are stored in the vectors $\boldsymbol{r}^o$ and $\boldsymbol{r}^p$, where the former contains the locations of the 121 training synoptic stations and the latter contains 37.207 centers of the grid cells of 4 km size over the area of Germany.

The next step is the attribution of a prior distribution for the model's parameters $[\boldsymbol{\beta}, \sigma^2, \phi]$. The specification of the prior is described in detail in appendix A. Once the observed data is prepared and the prior distributions specified, HIRAIN enters the step 1, where the BSGLM is run.

The model approximates the posterior distribution $[\boldsymbol{s}^o, \phi | \boldsymbol{y}^o, \boldsymbol{d}^o]$ through MCMC simulations and simulates $\boldsymbol{\beta}$ and σ^2 by sampling directly from the posterior distribution $[\boldsymbol{\beta}, \sigma^2 | s^o, \phi]$. Given the simulations of $\boldsymbol{s}^o$ and ϕ, $\boldsymbol{s}^p$ is simulated from the posterior predictive distribution $[\boldsymbol{s}^p | \boldsymbol{s}^o, \phi, \boldsymbol{d}^p] \sim t_{n+n_\sigma}(\boldsymbol{\mu}_1, S_1^2 \Sigma_1)$. This requires the inversion of the covariance matrix, which has dimenson of 37.207 x 37.207. A lot of computational capacity would be required. Thus, the area of Germany is divided in 8 smaller regions, in order to make the algorithm feasible.
In each region the simulations are conditioned on the borders of the neighborhood

Table 6.1.: Latitude and longitude delimiting the 8 areas in Germany.

Area	Longitude	Latitude
1	6° - 11°	52.125° - 55.000°
2	10° - 15°	52.125° - 55.000°
3	6° - 11°	50.300° - 52.725°
4	10° - 15°	50.300° - 52.725°
5	6° - 11°	48.875° - 50.900°
6	10° - 15°	48.875° - 50.900°
7	6° - 11°	47.400° - 49.475°
8	10° - 15°	47.400° - 49.475°

areas, which have 1° in latitude and 0.6° in longitude. This makes the covariance matrix much smaller and easier to be inverted. The 8 areas are shown by the map in the flow chart and the corresponding size is presented in table 6.1.

Within this approach the simulations are performed in the following sequence: s_j^p is simulated in the area 1 first. Given the simulations in the boundary region between area 2 and 1, samples are drawn for area 2. Simulations in the area 3 are conditioned on the simulations in the boundary region between area 1 and 3 and the region between area 2 and 3. The same logic is applied to the next areas until realisations of $\boldsymbol{s}^p$ are drawn at every location $\boldsymbol{r}^p$ in all 8 regions. Given the simulated values of $\boldsymbol{s}^p$, the model provides at the same locations of interest, $\boldsymbol{r}^p$, the probability of precipitation ≥ 0.1 mm/h, namely $\pi(\boldsymbol{r}^p)$, by applying the inverse logit function to the values of $\boldsymbol{s}^p$.

The end product of the BSGLM for time t_1 is a simulated spatial Gaussian process that drives occurrence/non-occurrence of precipitation ≥ 0.1 mm/h with 4km resolution and the corresponding probabilities of precipitation ≥ 0.1 mm/h.

The time is updated to $t = t + 1$, the corresponding input data is prepared and the BSGLM is run again. This procedure is repeated while $t \leq t_f$, where t_f is the last time of interest.

Once step 1 is completed, HIRAIN enters step 2 to generate occurrence/non-occurrence of precipitation ≥ 0.1 mm/h, namely $\boldsymbol{y}^p$. This is done through BNS and THR. In BNS realisations of 0's and 1's are drawn at each grid point r_j^p and time steps, independently. In THR occurrence/non-occurrence of precipitation ≥ 0.1 mm/h is obtained by thresholding the fields of $\pi(\boldsymbol{r}^p)$ with P_t.

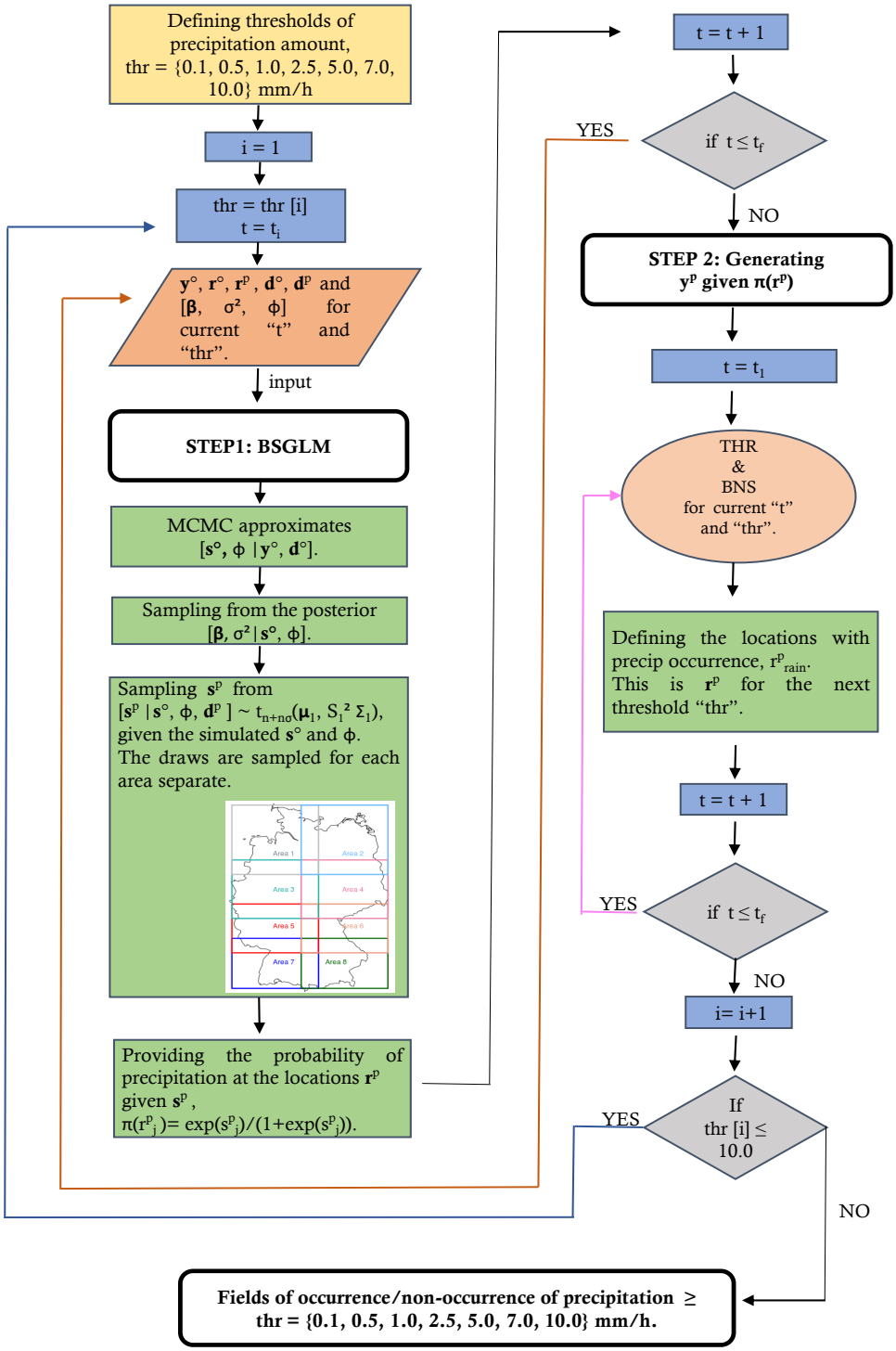

Figure 6.2.: Schematic representation of HIRAIN.

The product provided by the step 2 is the variable of interest, $\boldsymbol{y}^p$ with 4 km resolution from time t_1 to t_f.

After completing step 1 and 2, HIRAIN is applied to the next threshold of precipitation, namely 0.5 mm/h. The time is set to $t = t_1$ and the corresponding input data is prepared. Here values of $\boldsymbol{y}^o$ are obtained by assigning 1 to precipitation amount $\geq$ 0.5 mm/h and zero otherwise. The coordinates $\boldsymbol{r}^o$ and the covariate $\boldsymbol{d}^o$ do not change with the change of thr. However, the prediction locations are $\boldsymbol{r}^p = \boldsymbol{r}^p_{rain}$, where the vector $\boldsymbol{r}^p_{rain}$ contains the locations, for which precipitation $\geq$ 0.1 mm/h is simulated. The vector $\boldsymbol{d}^p$ contains the covariate at those locations. This procedure is adopted to ensure the generation of consistent precipitation product, although HIRAIN is applied to each threshold independently. With this approach precipitation $\geq$ 0.5 mm/h can only be predicted at locations where precipitation $\geq$ 0.1 mm/h is to be true. If there is no precipitation in the lower threshold, non-occurrence is assigned to the current threshold as well. Occurrence/non-occurrence of precipitation $\geq$ 0.5 mm/h with 4 km resolution is obtained for the time t_1 to t_f.

The threshold is then update to thr=1.0 mm/h and the same logic is used to generate the corresponding $\boldsymbol{y}^p$. HIRAIN is repeated for further thresholds following the same steps described above until thr=10.0 mm/h. Realisations of $\boldsymbol{y}^p$ for the different thresholds of precipitation amount are obtained as final product of HIRAIN.

7. Verification methods

This chapter presents the verification diagnostics used in this study to assess the performance of HIRAIN.

7.1. Brier skill score

The Brier skill score (*BSS*, Brier 1950) is used to measure the relative improvement of the probabilistic forecast over a reference *RS*

$$BSS = 1 - \frac{BS}{RS}, \tag{7.72}$$

where

$$BS = \frac{1}{l}\sum_{k=1}^{l}(p_k - o_k)^2. \tag{7.73}$$

BS is the Brier score and it compares the forecast probability p_k with the observation o_k, which has either a value of 1 or 0 whether the event occurred or not. The score is computed over the total number of events l. The climatology is usually used for the reference forecast *RS*. A perfect forecast provides *BSS* equal 1, negative values indicate that the model is worse than the reference forecast and a score equal 0 indicates no improvement in relation to the reference.

7.2. Ensemble Brier skill score

This study also uses the ensemble Brier skill score (*EBSS*) to assess the performance of HIRAIN. This skill score is an extention of the BSS to ensemble forecasts and it is based on the approach used by Schwartz *et al.* 2010.

By extending the *BSS* into ensemble forecasts, the forecast F for each ensemble

member is converted to binary information BP (0 or 1) based on the fact whether the forecast exceeds or not a threshold T

$$BP_{ki} = \begin{cases} 1 & \text{if } F_{ki} \geq T \\ 0 & \text{otherwise} \end{cases}$$

and thus an ensemble probability EP is computed by averaging BP over all members

$$EP_i = \frac{1}{n} \sum_{k=1}^{n} BP_{ki}, \tag{7.74}$$

where n is the total number of members, k refers to the kth ensemble member and i to the ith grid point. EP takes on values between 0 and 1.

In the same way the observations are converted to a binary information as well, namely BP_o. This is compared to the EP through the ensemble Brier score (EBS), which has the form of a Brier score and is written as

$$EBS = \frac{1}{N_d} \sum_{k=1}^{N_d} (EP_k - BP_{o_k})^2 \tag{7.75}$$

where N_d is the total number of grid points in the whole domain. The score can take on values between 0 and 1.

A skill score can then be computed in relation to a reference ERS. This is the $EBSS$ and it is given as

$$EBSS = 1 - \frac{EBS}{ERS}. \tag{7.76}$$

7.3. Fraction skill score

The fraction skill score (FSS) (Roberts, Lean 2008) is a version of the BSS in which over a neighbor region the frequency of the predicted events is compared to the frequency of the observations. "The purpose of this verification method is to obtain a measure of how forecast skill varies with spatial scale" (Roberts, Lean 2008). The idea behind it is that a model has good performance if the frequencies of predicted and observed events in the nearby area are similar (Ebert 1998).

In order to compute the FSS, prediction and observation fields are projected on to the same spatial grid. Appropriate thresholds T are chosen to convert predictions and observations to binary information, namely BP and BO, respectively. An one is assigned to the grid point k if the threshold is exceeded and zero otherwise,

$$BP_k = \begin{cases} 1 & \text{if } p_k \geq T \\ 0 & \text{otherwise} , \end{cases}$$

$$BO_k = \begin{cases} 1 & \text{if } o_k \geq T \\ 0 & \text{otherwise} , \end{cases}$$

where p_k and o_k is the prediction and the observation, respectively at the grid point k.

A spatial scale w is chosen to define a nearby area around each grid point, for which predictions and observations are compared. This scale indicates the number of nearby grid points that are used to delimit this area. Pointwise comparisons use a scale of w =0. This means that no nearby grid point is considered for the comparison. For instance if a scale of 2 is chosen, the area around each grid point is delimited by 2 grid points to the right, 2 grid points to the left, 2 grid points to the north and 2 grid points to the south. The same logic follows in the choice of other scales.

By averaging BP_k and BO_k over the defined area, the fractional coverage of predicted (NP) and observed (NO) events are obtained,

$$NP_i = \frac{1}{N} \sum_{k=1}^{N} BP_k \tag{7.77}$$

$$NO_i = \frac{1}{N} \sum_{k=1}^{N} BO_k, \tag{7.78}$$

where N is the total number of grid points in the area around i, and k refers to the kth grid point inside the region. Figure 7.1 shows a schematic representation of a nearby region surrounding the grid point represented by X with the corresponding fractional coverage of forecast and observed events.

NP can be compared to NO through a fraction Brier score (FBS),

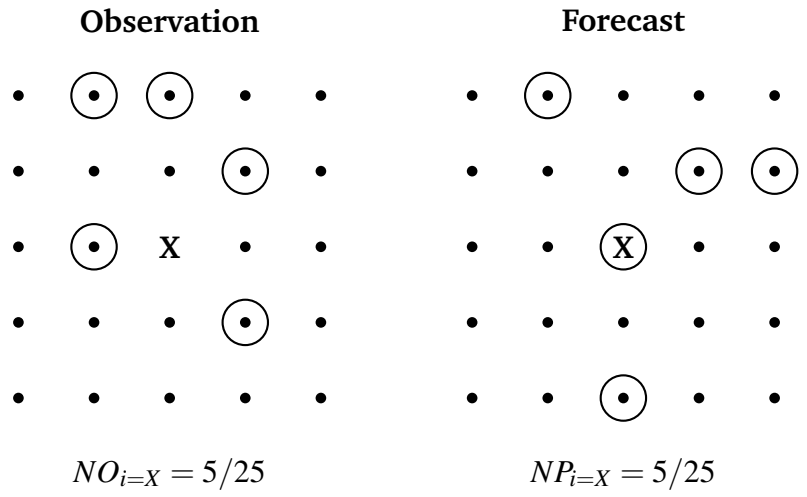

Figure 7.1.: Schematic representation of a nearby region surrounding the grid point **X** and the fractional coverage of events for the (left) observations and (rigt) model forecast. The grid points with a circle represent the locations with events.

$$FBS = \frac{1}{N_d} \sum_{i=1}^{N_d} (NP_i - NO_i)^2, \tag{7.79}$$

where N_d is the total number of grid points in the whole domain. The score can take on values between 0 and 1.

The FSS is computed with reference to the worst possible forecast when observed and forecast events do not overlap.

$$FSS = 1 - \frac{FBS}{\frac{1}{N_d}[\sum_{i=1}^{N_d} NP_i^2 + \sum_{i=1}^{N_d} NO_i^2]}. \tag{7.80}$$

This score can take on values between 0 (no overlap) and 1 (perfect match).

7.4. Categorical verification scores

Based on the joint distribution of predicted and observed events/non-events categorical verification scores are appropriate for validating binary responses. The joint distribution can be summarized by a categorical contingency table (Table 7.1) and the scores are computed based on its entries: hits, false alarms, misses and correct negatives.

Table 7.1.: Contingency table

	Observed		
Forecast	yes	no	total
yes	hits	false alarms	forecast yes
no	misses	correct negatives	forecast no
total	observed yes	observed no	total

The hits indicate the number of times that predicted events are observed. The false alarm gives the number of times that predicted events do not occur. Misses indicate how many times observed events are not predicted and the correct negatives are the number of times non-events are not predicted.

Based on this table, the verification scores used in this study are the following.

The frequency bias (BIAS),

$$BIAS = \frac{\text{hits} + \text{false alarms}}{\text{hits} + \text{misses}} \tag{7.81}$$

that provides the ratio between the number of forecast events and the observed events. The perfect score is 1 and BIAS can range from 0 to ∞. When BIAS <1 the model underestimates the frequency of events and for BIAS > 1 the frequency is overestimated.

The probability of detection (POD),

$$POD = \frac{\text{hits}}{\text{hits} \ + \ \text{misses}} \tag{7.82}$$

that provides the fraction of observed events that are correctly predicted. It ranges from 0 (observations and predictions do not overlap) to 1 (all the observed events are correct forecast).

The false alarm (FAR),

$$FAR = \frac{\text{false alarms}}{\text{hits} \ + \ \text{false alarms}} \tag{7.83}$$

that gives the fraction of predicted events that are not observed. The score ranges from 1 (no overlap between observations and predictions) to 0 (all predicted events are observed).

The odds ratio (OR),

$$OR = \frac{(\text{hits})(\text{correct negatives})}{(\text{misses})(\text{false alarms})} \tag{7.84}$$

that provides the ratio of the odds of a hit to the odds of a false alarm. Usually the logarithm of OR is used (Stephenson 2000). Higher log odds ratio indicate better performance. Log odds ratio equal 0 indicates no skill.

The equitable threat score (ETS), or also called Gilbert skill score (GSS),

$$ETS = \frac{\text{hits} - \text{hits random}}{\text{hits} + \text{misses} + \text{false alarms} - \text{hits random}}, \tag{7.85}$$

where

$$\text{hits random} = \frac{(\text{hits} + \text{misses})(\text{hits} + \text{false alarms})}{\text{hits} + \text{misses} + \text{false alarms} + \text{correct negatives}}, \tag{7.86}$$

that measures how well the predicted event corresponds to the observed event, accounting for hits due to random chance. It ranges from $\frac{-1}{3}$ to 1, where the latter is the best statistic. ETS equal zero indicates no skill.

The last statistic computed from the contingency table and used in this study is the Hanssen and Kuipers discriminant (HK), or also called true skill statistic (TSS) or Peirce's skill score (PSS),

$$HK = \frac{\text{hits}}{\text{hits} + \text{misses}} - \frac{\text{false alarms}}{\text{false alarms} + \text{correct negatives}}, \tag{7.87}$$

It measures how well the predictions are able to separate "yes" events from "no" events. It ranges from -1 to 1, where the latter is the best statistc. HK equal zero indicates no skill.

7.5. Lorelogram

Lorelogram is an alternative to the variogram when working with categorical data (Diggle 1990). It measures the dependence between correlated categorical responses. It was first used and proposed by Heagerty 1995 and it is based on marginal pairwise log odds ratio. Higher values of the lorelogram indicate stronger

dependence. In this study the lorelogram is computed in space. Thus, it typically decreases as the distance increases. This tool is used here to assess whether the generated fields are able to represent the spatial dependence structure of precipitation.

8. Probability of precipitation

This chapter presents the probability of precipitation $\pi(\boldsymbol{r}^p)$ obtained from the simulated $\boldsymbol{s}^p$ in the first step of HIRAIN. Results are shown here for precipitation $\geq$ 0.1mm/h. This threshold presents the highest frequency of occurrence among the thresholds throughout the year (Figure 8.2) and includes all possible intensities of precipitation, from light to heavy. As this is also the resolution with what the synoptic stations measure precipitation, it is the threshold that defines precipitation occurrence after all. Since the simulations for higher thresholds depend on the good performance of HIRAIN for the first threshold, the evaluation of the BSGLM to generate probability of precipitation $\geq$ 0.1 mm/h is shown here in more detail. The probabilities of precipitation to exceed further thresholds are discussed later in chapter 10.

8.1. Model evaluation

In step 1 of HIRAIN, the BSGLM is run for the hours with at least 3 training synoptic stations measuring precipitation $\geq$ 0.1 mm/h. It provides 10 realisations of $\boldsymbol{s}^p$ and the corresponding probabilities $\pi(\boldsymbol{r}^p)$. Missing values are assigned to these vectors for those hours with less than 3 stations with precipitation exceeding the threshold.

In order to assess the performance of the BSGLM, $\pi(\boldsymbol{r}^p)$ at 1384 validation stations is validated against the synoptic stations by means of the BSS. The climatology of these stations is used as reference and it is based on hourly precipitation data from 2004 to 2010. For the BSS, $\pi(\boldsymbol{r}^p)$ is computed using the median $\boldsymbol{s}^p$ of the posterior predictive distribution. For every hour with valid data a BSS is calculated. The box-and-whisker plot in Figure 8.1 displays the variation in the BSS within each month.

The full range of variation is shown by the whiskers representing the minimim

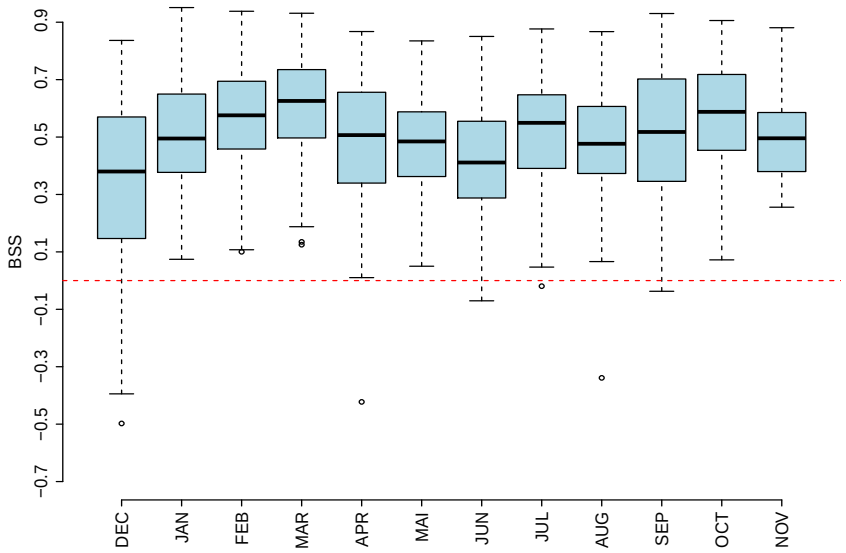

Figure 8.1.: BSS for the probability of precipitation $\geq$ 0.1mm/h. The score is computed hourly from December of 2010 to November of 2011. The dashed red line indicates no skill.

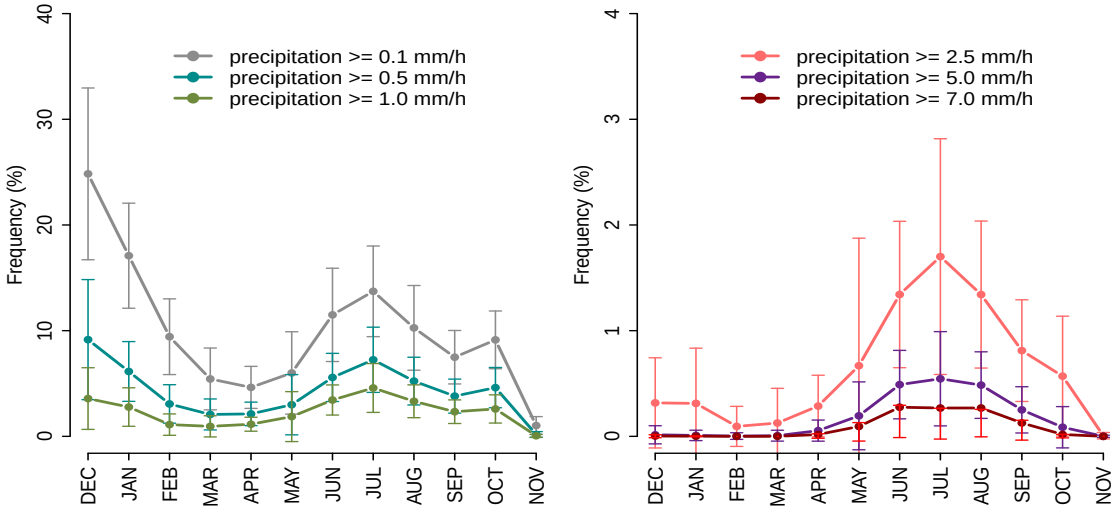

Figure 8.2.: Frequency of observed precipitation above a certain threshold at 1384 synoptic stations in the period between December of 2010 to November of 2011. The filled circles represent the mean frequency within a month and the whiskers the corresponding standard deviation. The graph on the left hand side has different y-axis than the graph on the right hand side.

and maximum BSS. The bottom and top of the boxes indicate the first (Q_1) and third (Q_3) quartile, respectively. The distance between them is the interquartile range (IQR) and provides a measure of the dispersion of the score within a month. The segment within the box is the median BSS and the open circles are the outliers. These are BSSs, which are either greater than $Q_3 + 1.5IQR$ or lower than $Q_1 - 1.5IQR$.

In every month there is an improvement in the BS with respect to the score for the climatological forecast. There is a skill in the predictions with respect to the climatology. Based on the median of each box, the BSS ranges mostly from 0.4 to 0.6. An annual cycle can also be identified, although there is no significant differences among the boxplots.

The skill score increases from December to March, decreases from April to the summer months and starts increasing again afterwards until November. From January to March the highest BSSs with the lowest dispersion around the median are found. During these months most of the precipitation events consisted of amounts not greater than 2.5 mm/h (Figure 8.2). October presents similar behaviour in the skill scores, however during this month precipitation $\geq$ 5.0 mm/h is also observed. Except for December of 2010, the lowest BSSs of the year are found in the late spring and in the begining of summer. During this period there is high frequency of precipitation with amounts ranging from 0.1 mm/h to amounts greater than 10 mm/h.

8.2. Categories of precipitation episodes

Three categories of precipitation cases with amount $\geq$ 0.1 mm/h are selected, in order to further investigate which are the factors that play major roles in the performance of the BSGLM and in the variability of the BSS throughout the year.

In these categories the following factors are considered for the investigation: a) spatial distribution of the synoptic stations with precipitation, b) relationship between observed precipitation and the covariate and c) the number of stations with precipitation. The categories are identified as category A, category B and category C.

In category A precipitation $\geq$ 0.1 mm/h is observed by 25 or more training sta

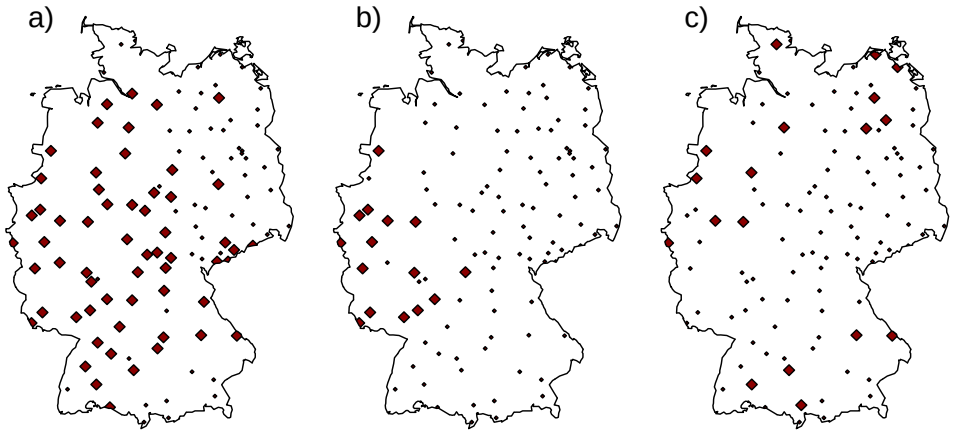

Figure 8.3.: Example of the spatial distribution of the synoptic training stations in categories (a) A, (b) B and (c) C.

tions, which are well distributed over the country. In category B, precipitation is observed by less than 25 stations, which are located over one region in Germany. Opposite to category B, in category C less than 25 synoptic stations observe precipitation, however they are wide distributed over the country.

An example of how the training stations are distributed in space within each category is shown in Figure 8.3. It is important to keep in mind that this is only an example. The stations do not need to be always located in western Germany in category B, for instance. They can be concentrated on any region of the country. The same is valid for categories A and C.

The precipitation cases studied in each category are visually selected. It is assured that there are 4 precipitation events per month in each category, which makes a total of 48 cases. Category A contains, however, only 46 cases, since there were very few precipitation events during November of 2011 that satisfied the requirements of category A. The selected cases are presented in table C.1 in the appendix C.

The BSS for the precipitation cases within each category is investigated (Figures 8.4, 8.5 and 8.6). In general the cases show variability in the skill score in every month. Although the boxplot in Figure 8.1 shows a tendency of the BSS to decrease as the summer approaches, this behaviour is not seen in the cases of the categories A, B and C. All three categories show that the skill score varies from

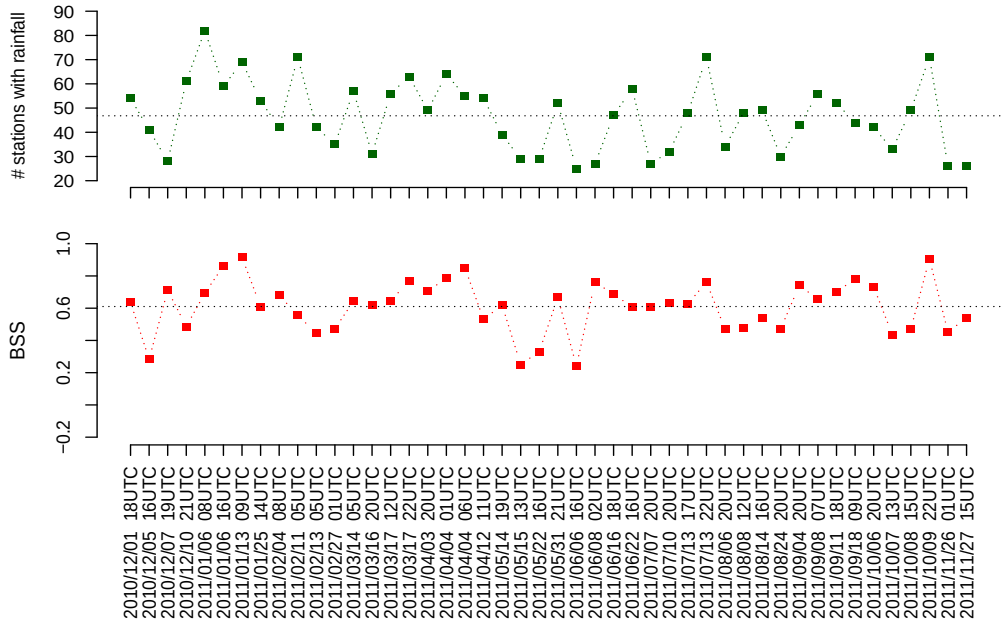

Figure 8.4.: (Top) Number of training synoptic stations with precipitation ≥ 0.1 mm/h for the cases in category A. The dashed black line represents the mean number of stations with precipitation. (Bottom) BSS for the probability of precipitation ≥ 0.1 mm/h. The dashed green line is the mean BSS.

low to high values in every month.

For instance, while on December 05th, 2011 at 16 UTC in category A the BSS is about 0.25, the next case on December 07th, 2011 at 19 UTC shows a much higher skill score of 0.7. During the summer on June 06th, 2011 at 16 UTC the BSS is about 0.22, whereas on June 08th, 2011 at 02 UTC the skill score is greater than 0.7. A similar behaviour is seen during the other seasons of the year and in the categories B and C as well.

On average, however, there is a difference in the BSS among the categories. The skill score averaged over the precipitation cases in category A equals to about 0.6. The same value is obtained in category B, whereas in category C the average is of about 0.4. Between category A and B there is not much difference in the skill scores of the corresponding precipitation cases. Both categories show scores ranging from 0.2 to 0.9. In category C, the BSS ranges from -0.15 to 0.71.

The studied cases indicate that the BSGLM performs well when precipitation is

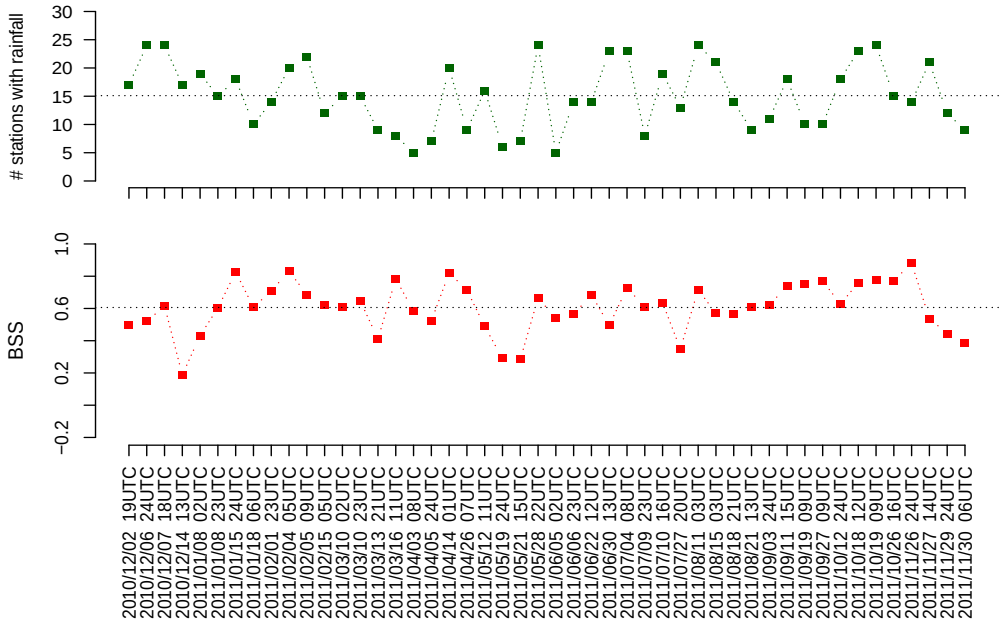

Figure 8.5.: Same as Fig. 8.4 but for precipitation cases in the category B.

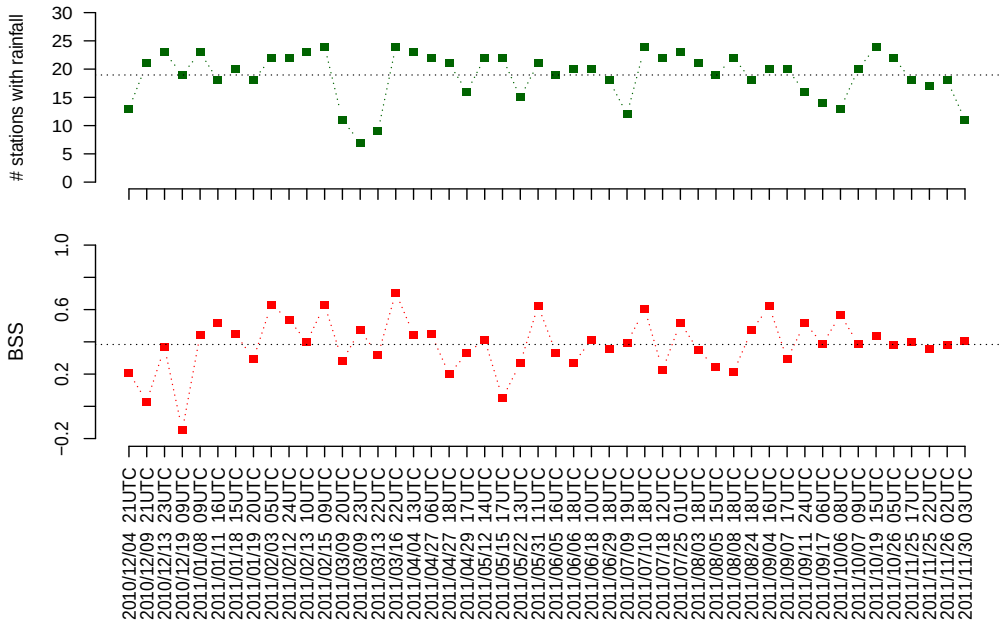

Figure 8.6.: Same as Fig. 8.4 but for precipitation cases in the category C.

observed by: a) more than 25 stations well distributed in space, and b) less than

25 stations located over one region in Germany. The performance of the model is less good when precipitation is observed by less than 25 stations that are wide distributed over the country.

8.3. Cases

In order to investigate further why the model performs differently for precipitation cases within the same category, two cases of each category are studied in more detail. These cases present high and low BSS, respectively. The realisations of $\pi(\boldsymbol{r}^p)$ are assessed and the correlation is computed between observed precipitation and IRBT at the training synoptic stations.

8.3.1. Category A

The precipitation events on May 15th, 2011 at 13 UTC and on June 22nd, 2011 at 16 UTC are selected within the category A. For the former the BSS is equal 0.25, whereas for the latter this score amounts 0.6. (Table 8.1).

During the precipitation event occurred on May 15th, 2011 at 13 UTC, 29 training synoptic stations distributed from northeastern to southwestern Germany measured precipitation amount $\geq$ 0.1 mm/h. This is shown by the big diamonds in Figure 8.7 (top left). The small diamonds represent the training stations with precipitation $<$ 0.1 mm/h. The IRBT is shown in colors in the same figure. The covariate ranges from 224.5 K to 283.3 K at the training synoptic stations. The remaining figures are three realisations of the probability of precipitation $\geq$ 0.1 mm/h simulated with the BSGLM. The validation synoptic stations with precipitation $\geq$ 0.1 mm/h ($<$ 0.1 mm/h) are shown in those same figures by the red (grey) dots.

The realisations show that the north and the southwest part of Germany are the regions with the highest probabilities of precipitation, with $\pi(\boldsymbol{r}^p) \geq 0.8$. Many of the validation synoptic stations with precipitation $\geq$ 0.1 mm/h (red dots) are, however, located outside these regions, where the simulated probability is $\pi(\boldsymbol{r}^p)$ $\leq$ 0.1. In a small area near the southwestern border of Germany, the validation stations do not show precipitation, although the probability is $\pi(\boldsymbol{r}^p) \geq 0.8$. The

Table 8.1.: Selected precipitation cases in category A and the corresponding BSS.

Cases	BSS
15. May 2011, 13 UTC	0.25
22. June 2011, 16 UTC	0.6

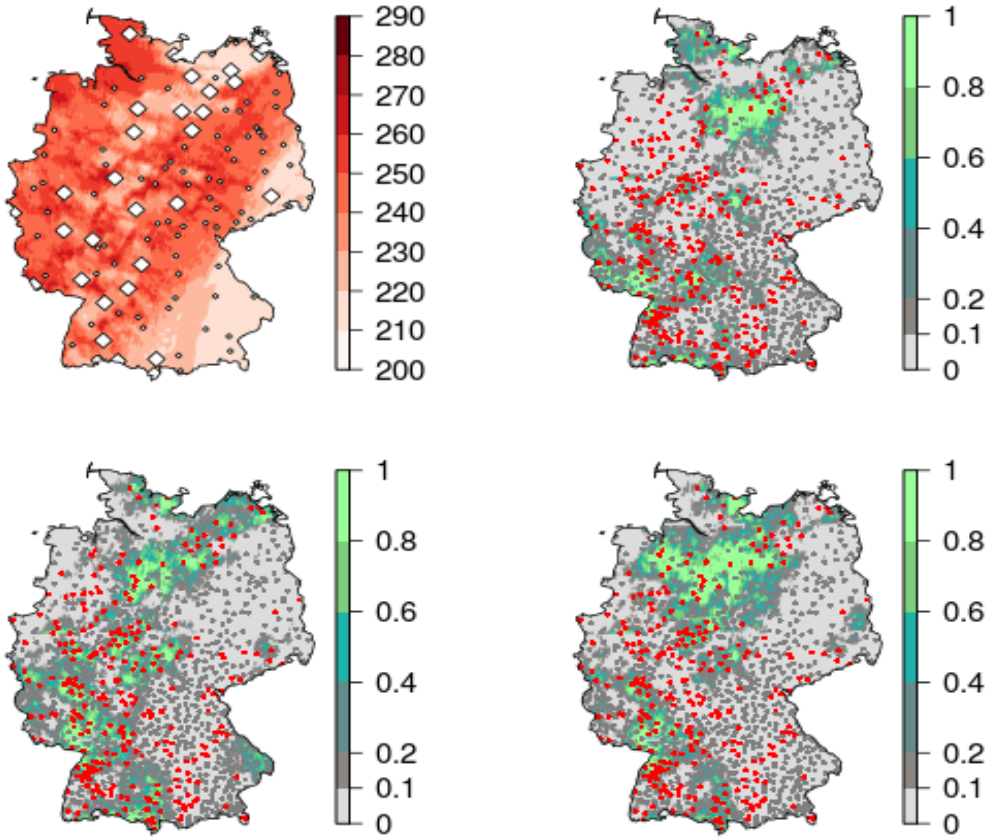

Figure 8.7.: (top left) IRBT on May 15th, 2011 at 13 UTC from MSG given in kelvins (in colors). The big (small) diamonds represent the training synoptic stations with precipitation $\geq$ 0.1 mm/h ($<$ 0.1 mm/h). The remaining maps are realisations of the probability of precipitation $\geq$ 0.1 mm/h. The red (grey) dots represent the validation synoptic stations with precipitation $\geq$ 0.1 mm/h ($<$ 0.1 mm/h).

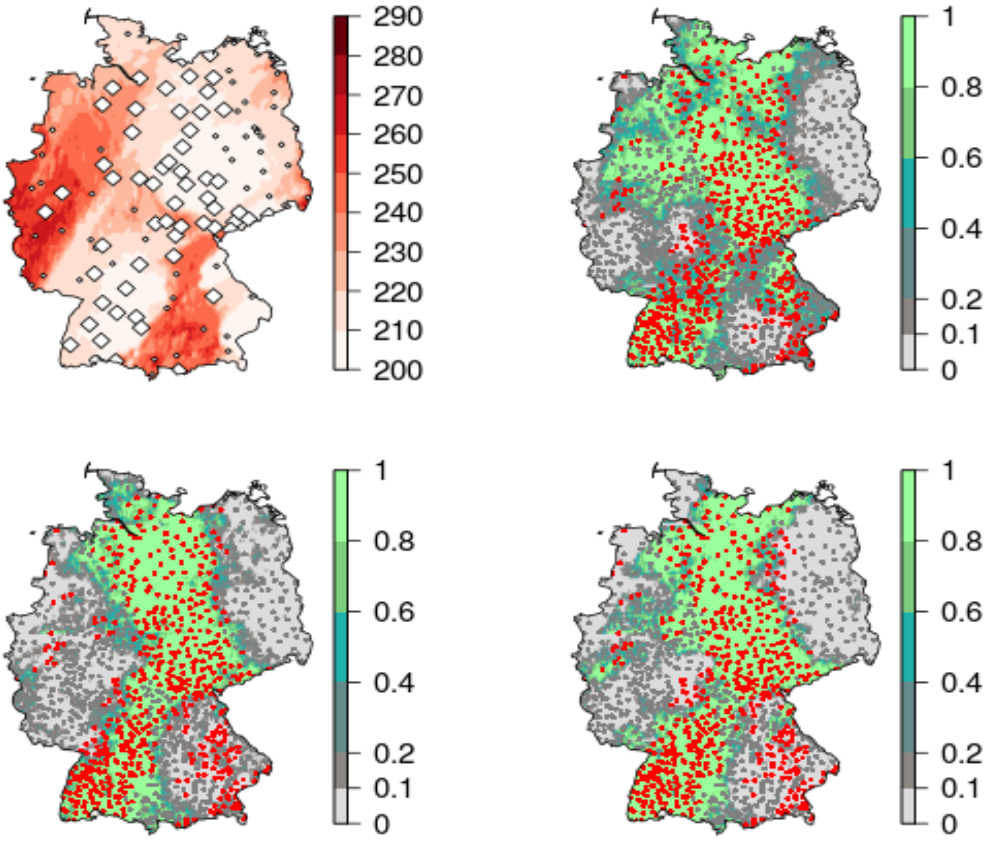

Figure 8.8.: Same as Fig. 8.7 but for precipitation on June 22nd, 2011 at 16 UTC.

disagreement between the validation stations and the predictions favours low BSS.

In order to investigate why the BSGLM performed poorly in this case, the correlation coefficient is computed between the precipitation amount at the training stations and the IRBT at the same locations. The idea is to verify whether the model's performance improves (worsens) given the agreement (disagreement) between the observations and the covariate. The coefficient and the summary of the significance test is presented in Appendix D.

The correlation coefficient is very low and equal -0.16. The correlation is expected to be negative since precipitation amount increases as the IRBT decreases and vice-versa. However, since the correlation is not significant at the 95% significance level, this association might also happen by chance in this case. Besides, the fact that the correlation is low indicates that low IRBT is not very often associated

with precipitation at the stations and vice-versa. The lack of agreement between station information and IRBT might influence the performance of the BSGLM.

For the second precipitation case selected in category A, namely the event occurred on June 22nd, 2011 at 16 UTC there are 58 training stations with precipitation amount ≥ 0.1 mm/h (Figure 8.8). The IRBT ranges from 213.1 K to 279.3 K at the training station locations.

The realisations show that most of the validation stations with precipitation $\geq$ 0.1 mm/h are located in regions with $\pi(\boldsymbol{r}^p) \geq 0.8$ and most of the stations with no precipitation are located in areas with $\pi(\boldsymbol{r}^p) \leq 0.1$. For a large number of locations, the simulated probability agrees well with the observations. This seems to favour a high BSS.

There are regions, however, where high probability of precipitation is simulated at locations where the validation stations do not observe precipitation. This is shown by the first realisation (Figure 8.8 (top right)). The probability of precipitation is $\pi(\boldsymbol{r}^p) \geq 0.8$ in northwestern Germany, while the validation stations do not measure precipitation. The opposite pattern is found in southeastern Germany, where very low probability ($\pi(\boldsymbol{r}^p) \leq 0.1$) is simulated, but the synoptic stations measured precipitation (Figures 8.8 (bottom)). For this area, the covariate provides low values of IRBT, but there is a very reduced number of training stations and most of them do not measure precipitation. The fact, that the validation stations are not in good agreement with the predicted probabilities in this area does not mean that the model has bad performance, but that the data obtained from the training stations do not provide enough information for the region. The disagreement in those regions might influence the value of the BSS.

In both areas, the northwest and southeast of Germany, where the simulations do not agree with the validation stations are the areas where the observations and the covariate do not agree. In the wide region extending from the north to the very southweast there is very good agreement between the simulations and the observations. This is also the area for which observed precipitation is associated with very low IRBT. This is confirmed by the significant correlation coefficient of -0.3 between precipitation and IRBT.

Table 8.2.: Selected precipitation cases in category B and the corresponding BSS.

Cases	BSS
14. December 2010, 13 UTC	0.19
04. July 2011, 08 UTC	0.73

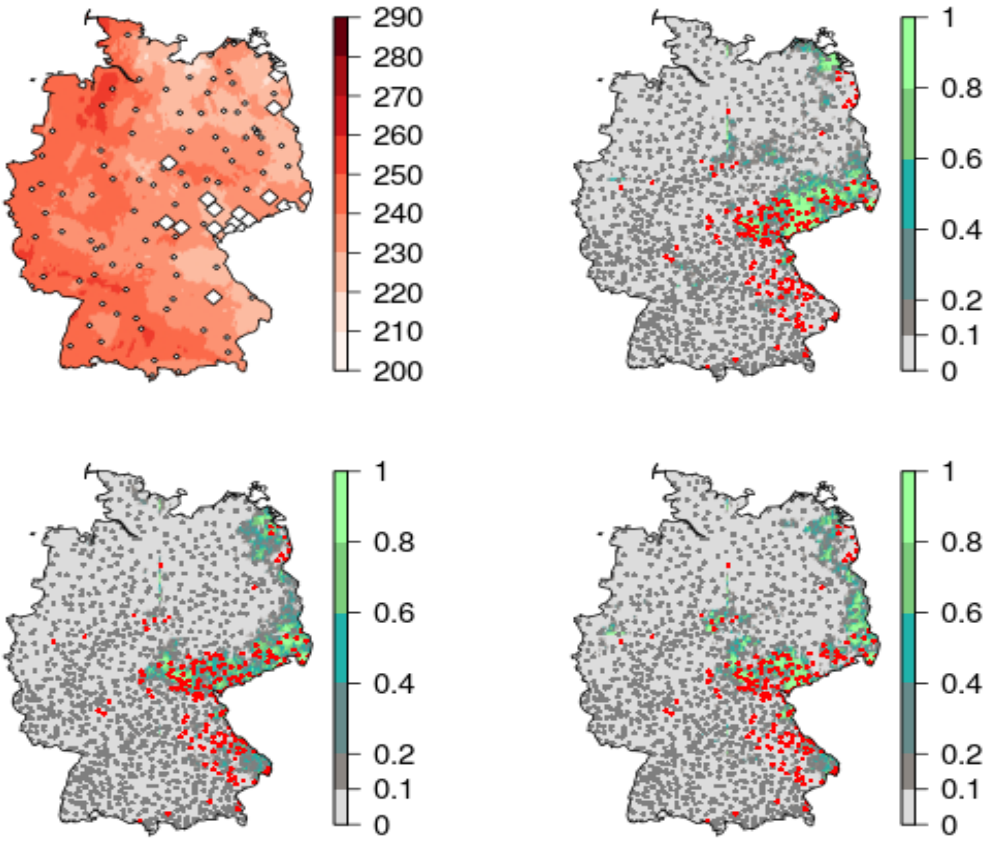

Figure 8.9.: Same as Fig. 8.7 but for precipitation on December 14th, 2010 at 13 UTC.

8.3.2. Category B

The precipitation cases selected in this category occurred on December 14th, 2010 at 13 UTC and on July 04th, 2011 at 08 UTC. The BSS obtained for these events amounts to 0.19 and 0.73, respectively (table 8.2).

On December 14th, 2010 at 13 UTC there are 17 training stations with precipitation $\geq$ 0.1 mm/h. The IRBT at the station locations ranges from 230.1 K to

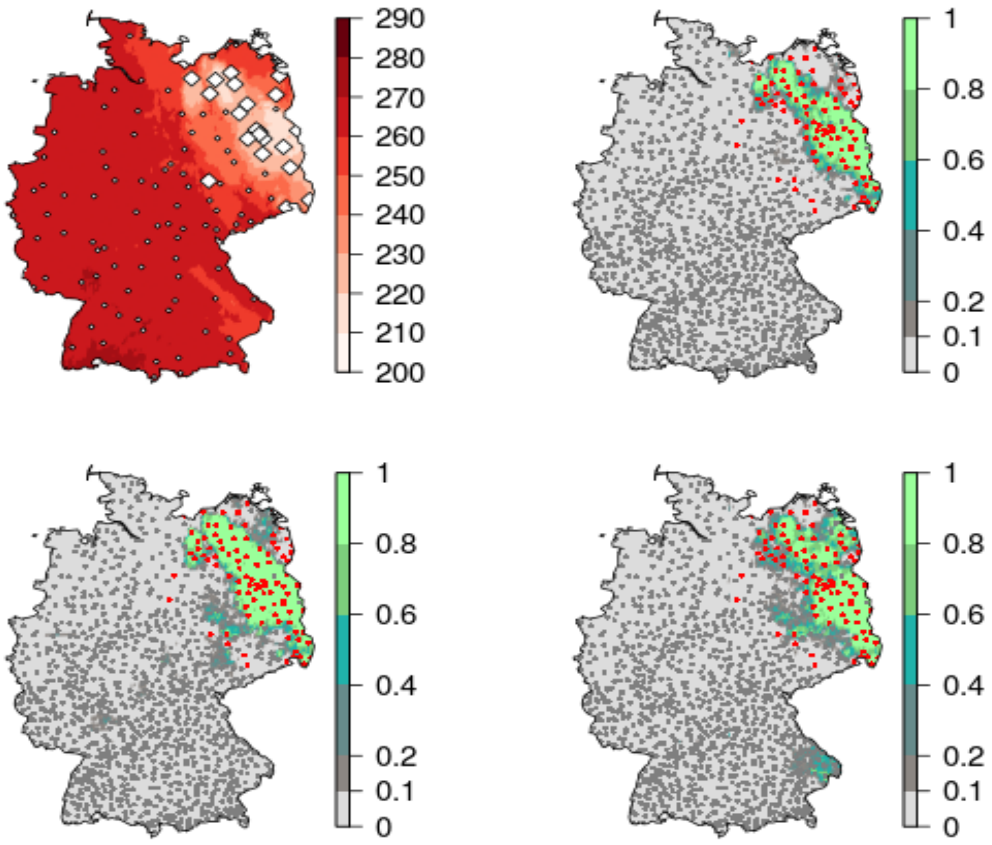

Figure 8.10.: Same as Fig. 8.7 but for precipitation on July 04th, 2011 at 08 UTC.

267.8 K (Figure 8.9). Both realisations in the bottom of Figure 8.9 show probability $\pi(\boldsymbol{r}^p) \geq 0.8$ in the border region of the northeast area. However, most of the validation stations in this region do not show precipitation occurrence.

All three realisations show very low probability of precipitation in the southeast region, where many validation stations observed precipitation. The simulations agree better with the validation stations in the central-east area, where there is a greater aggregation of training synoptic stations. The northeast and southeast area are not very well covered by training stations. Hence, not sufficient information from observations are provided for that region. Another factor that might influence the performance of the model is the absence of significant correlation between precipitation and IRBT (see Appendix D). Beside that, the fact that ob-

servations from outside the country are not used in the model might play a role in the simulations as well.

For the case on July 04th, 2011 at 08 UTC, 23 training stations observed precipitation $\geq$ 0.1 mm/h and the IRBT at the observation locations ranges from 226.6 K to 289.0 K (Figure 8.10 (top left)). The observations and IRBT are in very good agreement with a significant correlation of -0.67. The high coefficient indicates that precipitation is associated with low IRBT. The average IRBT found at locations with precipitation $\geq$ 0.1mm/h is about 245.3 K and for stations without precipitation the IRBT shows higher mean of about 277.6 K. The good agreement between observed precipitation and the covariate seems to favour the good performance of the model. Probability $\pi(\boldsymbol{r}^p) \geq 0.8$ is simulated in the northeast region, where precipitation is observed by the synoptic stations. In the areas without precipitation the probability is $\pi(\boldsymbol{r}^p) \leq 0.1$.

8.3.3. Category C

In this category precipitation on June 05th, 2011 at 16 UTC and on October 19th, 2011 at 15 UTC is investigated in more detail. For the former the BSS equals 0.33 and for the latter the skill score is 0.44 (table 8.3).

On June 5th, 2011 at 16 UTC, 19 training stations observed precipitation $\geq$ 0.1mm/h and the IRBT at the observation locations ranges from 207.8 K to 294.9 K. At the station locations with precipitation the average IRBT is about 232.9 K, whereas at the remaining stations this amounts 263.4 K. (Figure 8.11 (top left)). The correlation coefficient is significant and equal -0.24. However, the fact that the value is low indicates that precipitation occurrence is not always associated with low IRBT at the training station locations.

HIRAIN generated probabilities $\pi(\boldsymbol{r}^p) \geq 0.8$ for the majority of the areas where the validation stations detected precipitation. However, there are regions where this is not the case, as in the central-east area, where some realisations simulated probabilities $\pi(\boldsymbol{r}^p) \leq 0.1$, while the validation stations observed precipitation. This area is not well covered by the training stations (Figure 8.11 (top left)). Besides, the large distance among them in the entire region going from western to the central-eastern Germany does not favour the predictions. The region in the west part of Germany presents prob ability $\pi(\boldsymbol{r}^p) \geq 0.8$, while the validation stations do

Table 8.3.: Selected precipitation cases in category C and the corresponding BSS.

Cases	BSS
05. June 2011, 16 UTC	0.33
19. October 2011, 15 UTC	0.44

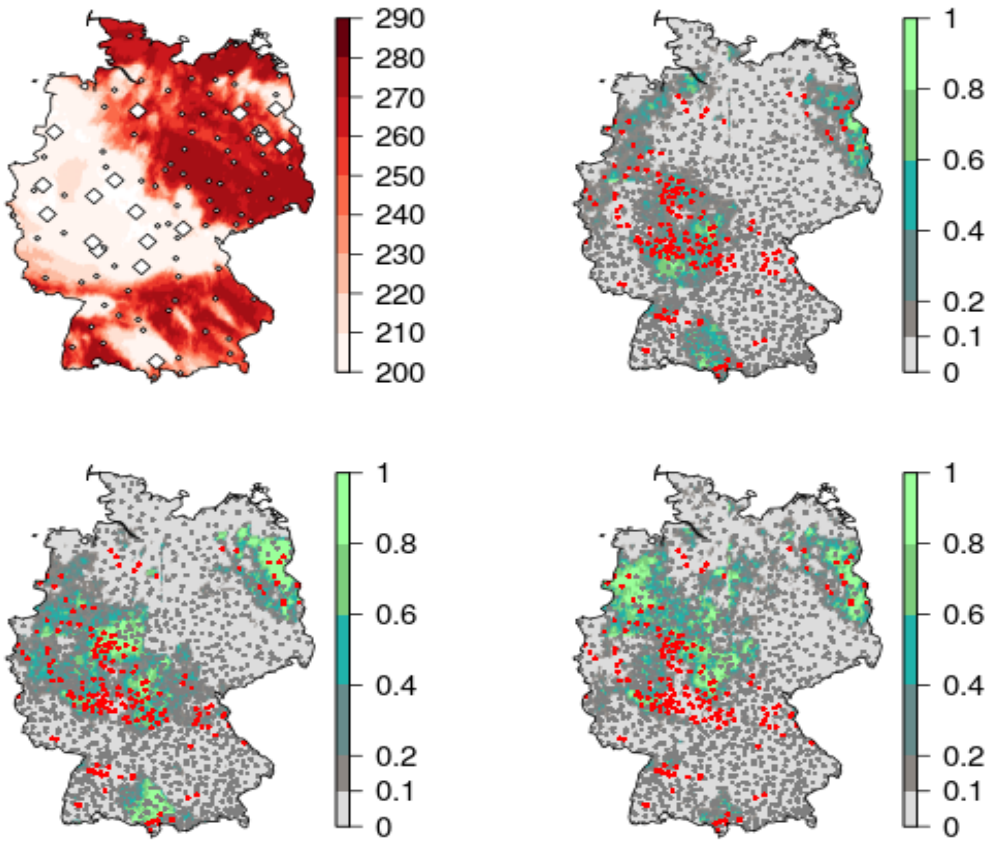

Figure 8.11.: Same as Fig. 8.7 but for precipitation on June 05th, 2011 at 16 UTC.

not observe precipitation. A denser network of training stations might improve predictions in this case.

For the precipitation case on October 19th, 2011 at 15 UTC, 24 stations measured precipitation and the covariate at the observation locations ranges from 227.2 K to 277.3 K (Figure 8.12 (top left)). The agreement between the probabili-

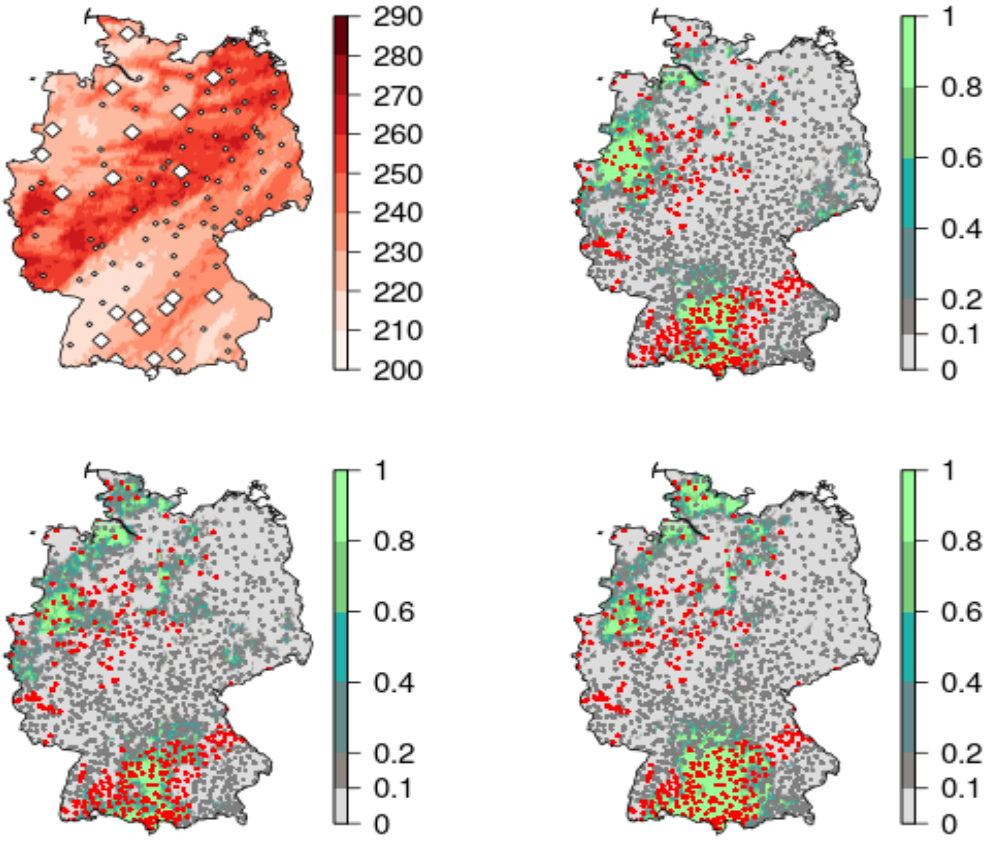

Figure 8.12.: Same as Fig. 8.7 but for precipitation on October 19th, 2011 at 15 UTC.

ties and the observations at the validation stations is better in southern than in northwestern Germany. In the northwest, although observations and probabilities agree, there are also validation stations with precipitation outside the area with probability $\pi(\boldsymbol{r}^p) \geq 0.8$. Instead, they are located in areas with probability $\pi(\boldsymbol{r}^p) \leq 0.1$. The lack of significant correlation between precipitation and IRBT (Appendix D) might also contribute to decrease the performance of the model in some areas.

From the cases investigated in all three categories, the results suggest that the performance of HIRAIN is influenced by the number of stations with precipitation ≥ 0.1 mm/h, their spatial distribution and by the relationship between precipitation and covariate.

The more negative the correlation coefficient is between precipitation and IRBT, the better are the predictions with the BSGLM and vice-versa. Beside that, the

greater the number of training stations with precipitation and the shorter the distance among them in space, the better the performance of the model.

During DJF the precipitation is mostly originated from stratiform cloud systems. These are usually well organized and present larger extensions than the convective precipitating systems. Thus, stratiform precipitation is often well distributed spatially. These facts associated with low IRBT comprehend the conditions found in category A that leads to good performance of the model.

During other seasons, specially during summer, the number of convective precipitation systems increases. Such precipitation events are usually more local and associated with low IRBT, which are the characteristics found in category B that also favour good performance of the BSGLM. However, the model will perform as in category C in the presence of precipitation at only a few locations with small spatial coverage and large geographical distance among them. In this scenario predictions tend to be less good than when characteristics of category A and B are present.

The fact that high and low BSS are found in the categories A, B and C (Figures 8.4 to 8.6) and that the BSGLM presents better performance during winter than during the other seasons (Figure 8.1) indicates that favorable conditions for high performance of the BSGLM appear more often during winter. The tendency of the BSS to decrease in the summer months suggests that unfavorable conditions appear more frequently in that period than during other seasons.

9. Occurrence/non-occurrence of precipitation

This chapter presents the results obtained from step 2 of HIRAIN, namely the occurrence/non-occurrence of precipitation $\mathbf{y}^p$. The focus here is on the fields generated for precipitation $\geq$ 0.1 mm/h. Fields for further thresholds will be presented in chapter 10.

Occurrence/non-occurrence of precipitation is generated through BNS and THR. Since THR requires a probability threshold (P_t) to produce the fields, the chapter starts showing how to determine P_t. This is followed by the assessment of the HIRAIN's performance by validating $\boldsymbol{y}^p$ from BNS and THR, respectively against radar estimates.

9.1. Probability threshold P_t

Realisations of probability of precipitation $\geq$ 0.1 mm/h are validated against the validation synoptic stations (Figure 3.2 (right)), in order to obtain P_t.

To this end, the realisations of $\pi(\boldsymbol{r}^p)$ are converted to binary fields. First the probability of 0.1 is used as a threshold. One is assigned to the grid points where $\pi(\boldsymbol{r}^p)$ is $\geq$ 0.1 and zero elsewhere. From the binary fields, an EP is computed by averaging the 10 realisations at every hour (equation 7.74). Precipitation amount of the validation stations are also converted to binary information, namely BP_o. One is assigned to stations with precipitation $\geq$ 0.1 mm/h and zero elsewhere. Lastly, BP_o is compared to EP through the computation of an EBS (equation 7.75).

The EBS is computed at every hour with valid data in the period from December of 2010 to November of 2011. The same logic is followed for the other probabilities from 0.2 up to 0.9. The scores are shown by the box-and-whisker plot in

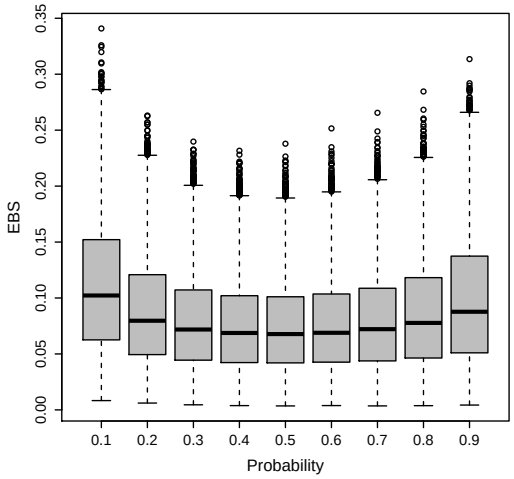

Figure 9.1.: EBS for precipitation ≥ 0.1 mm/h. Probabilities ranging between 0.1 and 0.9 are used as threshold to define occurrence/non-occurrence of precipitation. EBS is computed at every hour with valid data in the period from December of 2010 to November of 2011.

Figure 9.1. This shows how the EBS varies for each probability used as a threshold to define occurrence/non-occurrence of precipitation. The whiskers show the minimim and maximum EBS. The bottom and top of the boxes indicate the quartiles Q_1 and Q_3, respectively. The segment within the box is the median EBS and the open circles are the outliers.

The probabilities between 0.4 and 0.5 are the best to define occurrence/non-occurrence of precipitation ≥ 0.1 mm/h. Although the differences among the EBSs are not significant, the score is the closest to zero and has the lowest uncertainties for the probabilities 0.4 and 0.5.

A contingency table is computed, in order to further investigate which probability is the best to be used as threshold to generate $\boldsymbol{y}^p$ with THR. As before, binary fields are produced by thresholding $\pi(\boldsymbol{r}^p)$ using the probabilities from 0.1 to 0.9. These fields are compared to the observed binary information from the validation stations. A separate contingency table is set up for each probability used to generate the binary fields. The categorical statistics POD, FAR, BIAS and log odds ratio

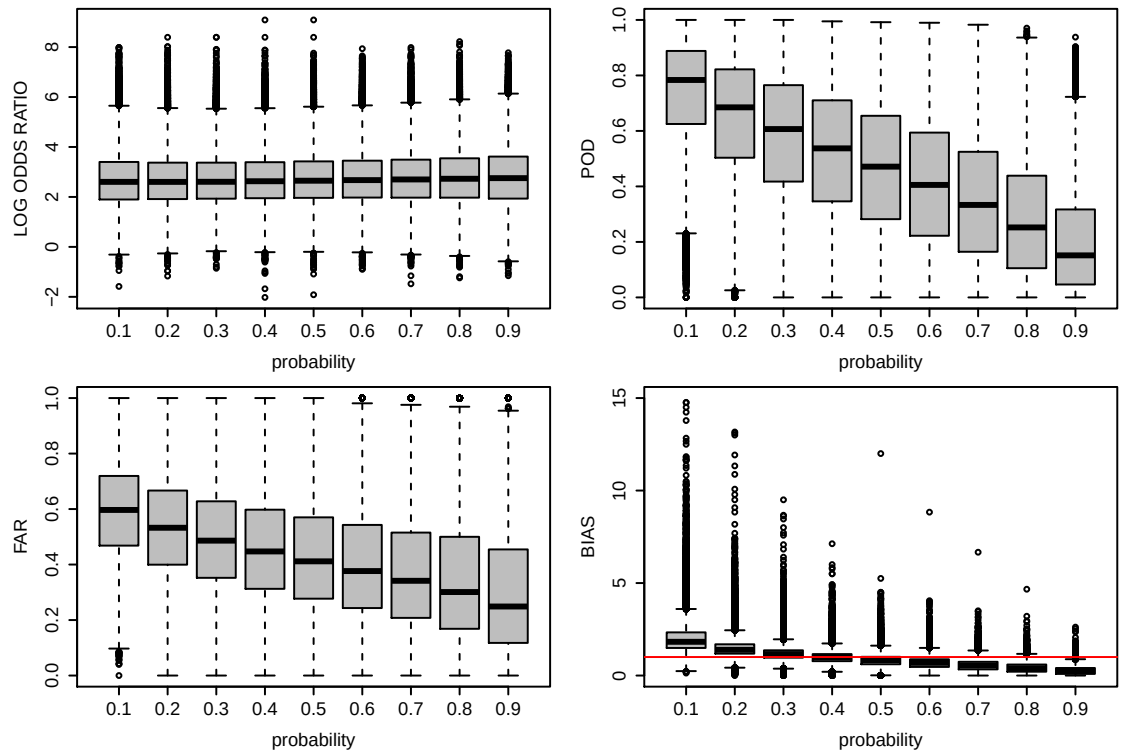

Figure 9.2.: Log odds ratio, POD, FAR and BIAS for precipitation $\geq$ 0.1 mm/h. Probabilities ranging between 0.1 and 0.9 are used as threshold to define occurrence/non-occurrence of precipitation. The categorical statistics are computed at every hour with valid data in the period from December of 2010 to November of 2011. The red line in the bottom right figure indicates BIAS equal 1.

are computed from these tables at every hour with valid data. Figure 9.2 shows the boxplots of them.

The Log odds ratio shows a median of approximately 2.5 for all probabilities. The fact that this statistic has value greater than zero indicates that the number of correct predictions is greater than the number of false alarms. The ratio does not change with the probability used as threshold. The number of correct detections increases on the costs of the increase in false alarms when the probability varies from 0.1 to 0.9. This statistic alone do not provide valuable information for the selection of P_t. The remaining statistics (POD, FAR and BIAS), however, change significantly with the probability. Their values decrease as the probability increases.

The POD is the best for the probability of 0.1 (median equal 0.8), however, on

the cost of a high value of FAR (median of about 0.6). The opposite is seen for the probability of 0.9, for which the best FAR (with median around 0.25) is given on the cost of the worst POD (with median of 0.15). The most appropriate probability to generate $\boldsymbol{y}^p$ with THR should provide the highest possible POD with the lowest possible FAR.

BIAS values show that probabilities greater than 0.4 used as threshold underestimate the number of precipitation occurrences in comparison with the observations. Probabilities lower than 0.3, however, overestimate them. Probabilities between 0.3 and 0.4 are most appropriate to detect precipitation occurrences, since the median of the BIAS is close to 1.

From the joint analysis of the EBS, POD, FAR and BIAS, the probability of 0.4 is selected for P_t. This provides the best statistics when occurrence/non-occurrence of precipitation $\geq$ 0.1 mm/h is generated with THR.

9.2. Validation against meteorological radar

Since the final product obtained by HIRAIN aims at replacing radar estimates in the assimilation scheme of the COSMO model, the fields of $\boldsymbol{y}^p$ are compared with radar estimates. It is important, however, to keep in mind that radar does not provide observations but remotely-sensed estimates. On one hand, error sources such as the attenuation of the radar beam, beam blockage and clutter may introduce large uncertainties in the precipitation estimates. On the other hand, the spatial coverage of radar allows for a better estimation of the spatial structure and dependence of precipitation than raingauges do.

Hourly radar precipitation is aggregated on a grid with 4 km spatial resolution and coverted to a binary field. An one is assigned to the grid points with precipitation $\geq$ 0.1 mm/h and zero elsewhere. Fields of $\boldsymbol{y}^p$ obtained through THR are produced by thresholding $\pi(\boldsymbol{r}^p)$ at the above defined value of P_t=0.4. The BNS samples from the Bernoulli distribution given $\pi(\boldsymbol{r}^p)$ at each grid point to provide $\boldsymbol{y}^p$.

The realisations from HIRAIN are compared to radar for the precipitation cases in the categories A, B and C. A contingency table is set up for comparison of THR and BNS with radar, respectively. The categorical statistics BIAS, POD, FAR, log

odds ratio, ETS and the HK are computed from these tables. It is also investigated whether HIRAIN is able to reproduce the spatial dependence structure of precipitation shown by radar.

9.2.1. Category A

9.2.1.1. Categorical statistics

In general THR leads to higher BIAS, POD, log odds ratio, ETS and HK than BNS (Figures 9.3 and 9.4). However, THR and BNS produce similar values of FAR.

For most cases BIAS varies between 0.5 and 1.5 for both methods. The POD ranges between 0.3 and 0.9, indicating that 30% to 90% of the occurrence of precipitation $\geq$ 0.1 mm/h is detected by the realisations. This is obtained on the costs of a FAR that ranges mostly between 0.1 and 0.5, i.e., 10% to 50% of the predicted occurrence of precipitation is not observed by radar.

The log odds ratio, ETS and HK are greater than zero for all precipitation events in the category. The former indicates that the odds of making a hit is greater than the odds of making a false alarm. The ETS provides how well the predicted occurrence of precipitation corresponds to the observation adjusted for hits to happen by chance. This reaches up to 0.3. The HK indicates that the predictions are able to separate between occurrence and non-occurrence of precipitation and reaches up to 0.4.

In order to exemplify the results shown by the categorical statistics, a few interesting cases will be described below.

On November 27th, 2011 at 15 UTC the realisations of $\mathbf{y}^p$ produced the highest BIAS in the category A (BIAS about 2.5). The POD is approximately 0.8 on the cost of a FAR that amounts to 0.7. The ability to predict precipitation occurrence, accounting for hits to happen by chance produced an ETS < 0.1, while the ability to separate between occurrence and non-occurrence of precipitation shows a HK of about 0.15. During this event, the training synoptic stations measured occurrence of precipitation $\geq$ 0.1 mm/h at many more locations than radar (not shown).

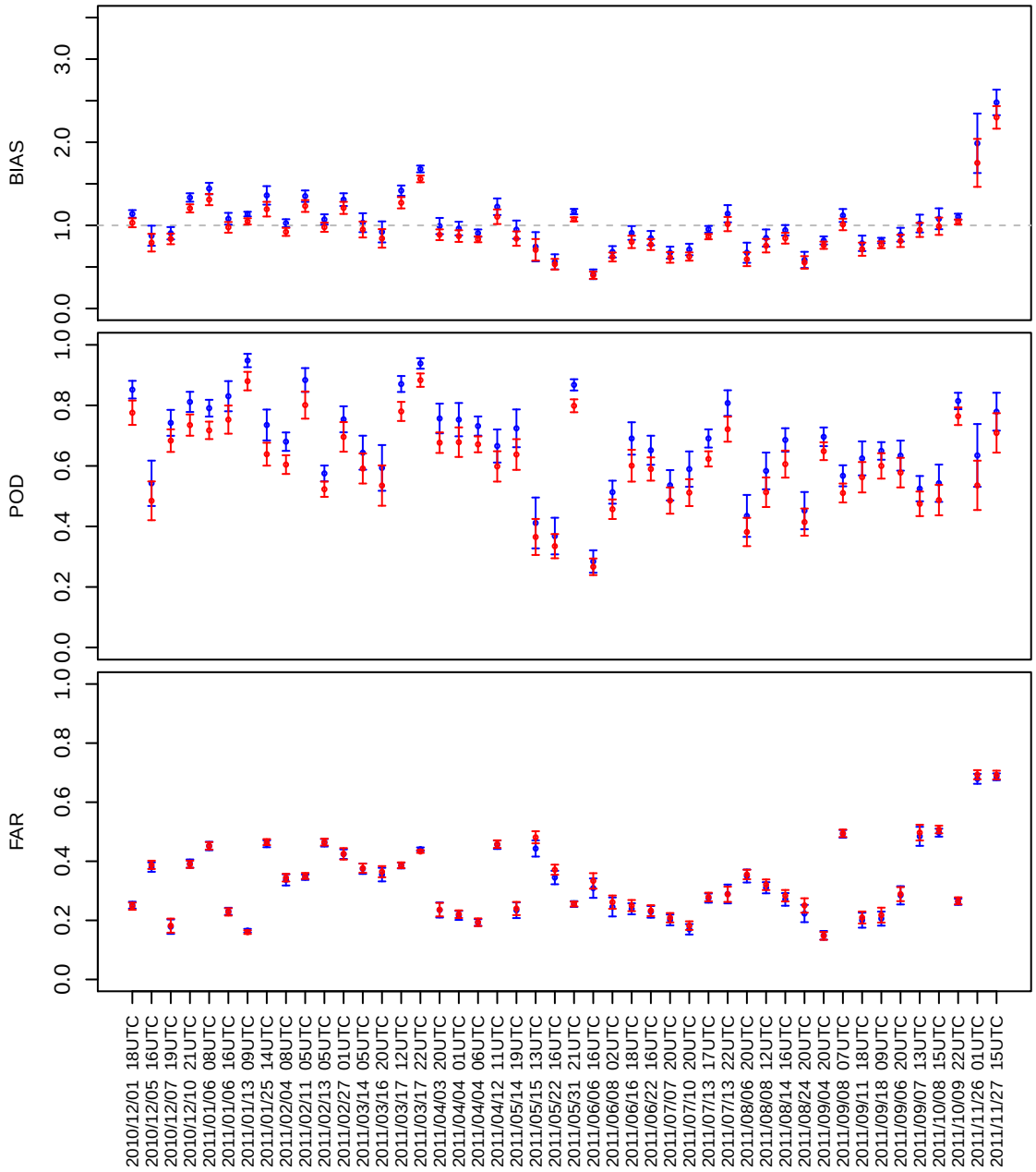

Figure 9.3.: BIAS, POD and FAR for the precipitation cases in category A. The categorical statistics are computed by comparing radar occurrence/non-occurrence of precipitation ≥ 0.1 mm/h to 10 realisations obtained with THR (blue) and to 10 realisations from BNS (red). The filled squares represent the average over 10 realisations and the whiskers, the standard deviation.

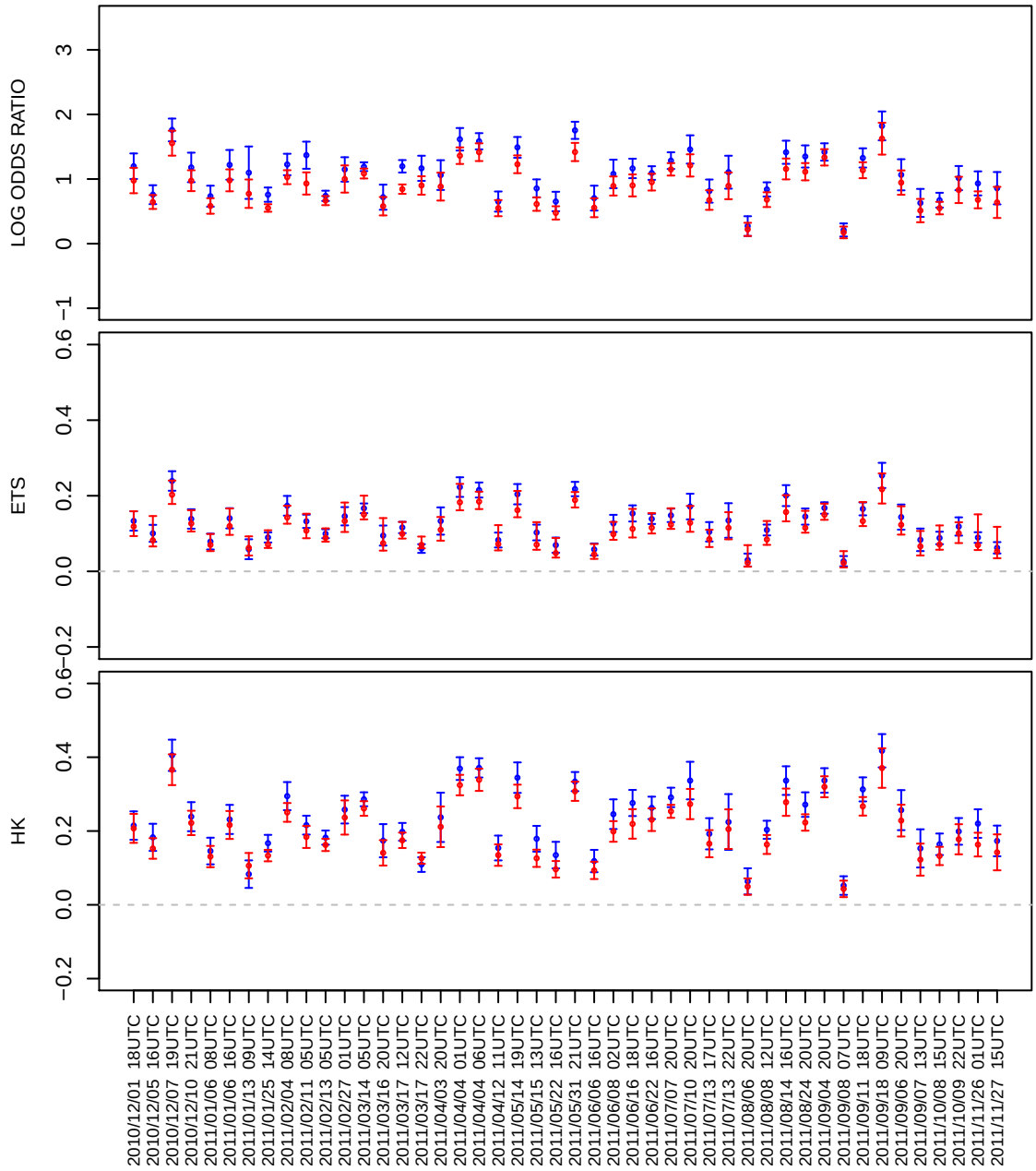

Figure 9.4.: Same as Fig. 9.3 but for the categorical statistics log odds ratio, ETS and HK.

Since the predictions are conditioned on the synoptic stations only and do not include radar information, the generated fields of $\boldsymbol{y}^p$ agree better with the stations than with radar. Precipitation occurrence is overestimated in comparison to radar

not because of bad performance of the model, but because the observations from stations and radar do not agree.

The precipitation event on May 31st, 2011 at 21 UTC shows an example with good results for all categorical statistics. BIAS is of about 1 for THR and BNS. The realisations of $\mathbf{y}^p$ predict about 80% (THR) and 90% (BNS) of the precipitation occurrences observed by radar. The false alarms produced by both methods amounts to 20%. The log odds ratio is greater than zero, amounting to 1.7 for THR and 1.3 for BNS. The ETS and HK amounts to 0.2 and 0.3, respectively. For this event, radar and synoptic stations observed precipitation $\geq$ 0.1 mm/h in the same region, which comprehends a large area in southern Germany and a smaller region in the northeast part of the country (not shown). The covariate also agrees well with the observed precipitation.

The precipitation event on June 06th, 2011 at 16UTC shows a case for which all categorical statistics presented less good results. BIAS is approximately 0.5, the realisations predict only 30% of the observed precipitation occurrences on the cost of about 30% false alarm. Log odds ratio is approximately 0.6 and there is a slight skill for correct predictions, accounting for hits to happen by chance (ETS $\sim$ 0.08). The realisations present also weak ability to separate between occurrence and non-occurrence of precipitation $\geq$ 0.1 mm/h. This is indicated by the low value of HK of approximately 0.1. During this event precipitation $\geq$ 0.1 mm/h is distributed over most part of the country. However, radar observed precipitation over larger areas than the synoptic stations (not shown). The covariate shows low IRBT in larger areas than the rainy areas observed by the stations.

9.2.1.2. Spatial dependence structure

Fields of occurrence/non-occurrence of precipitation $\geq$ 0.1 mm/h generated with THR and BNS are presented for the cases on June 22nd, 2011 at 16 UTC and on May 15th, 2011 at 13 UTC (selected in chapter 8). The case in June is shown here, whereas the event occurred in May is shown in appendix D.

Table 9.1 summarizes the categorical statistics for the event in June and Figure 9.5 shows the area with precipitation observed by radar (in shaded) and two realisations obtained with THR and BNS, respectively.

Table 9.1.: Categorical statistics for the precipitation case on June 22nd, 2011 at 16 UTC.

Statistics	THR	BNS
BIAS	0.85	0.77
POD	0.65	0.59
FAR	0.23	0.23
Log OR	1.09	0.95
ETS	0.14	0.11
HK	0.26	0.23

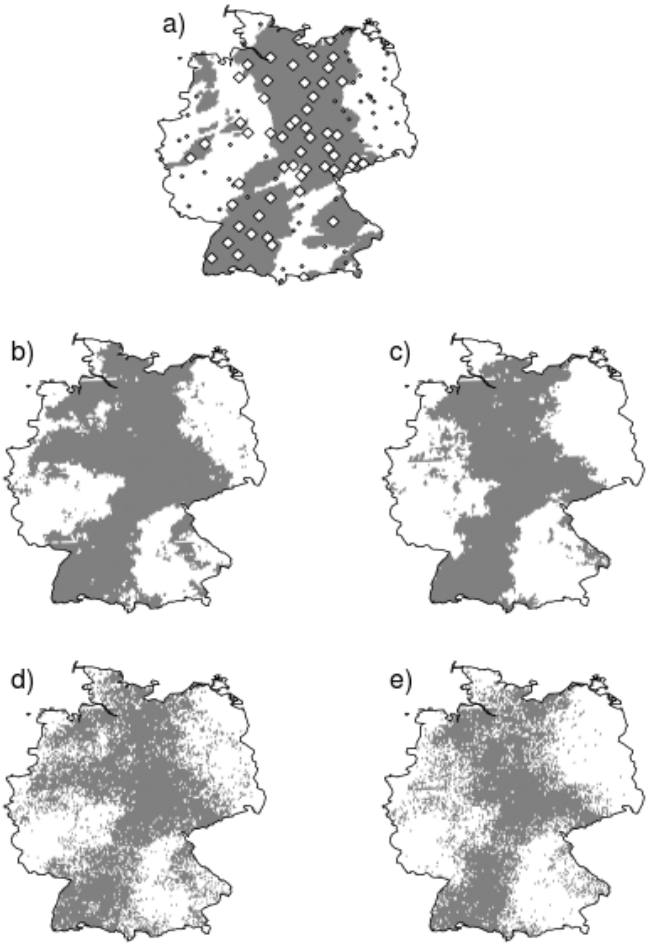

Figure 9.5.: a) Occurrence/non-occurrence of precipitation $\geq$ 0.1 mm/h on June 22nd, 2011 at 16 UTC observed by meteorological radar. Figures b) and c) show two realisations obtained with THR and figures d) and e) present two realisations from BNS.

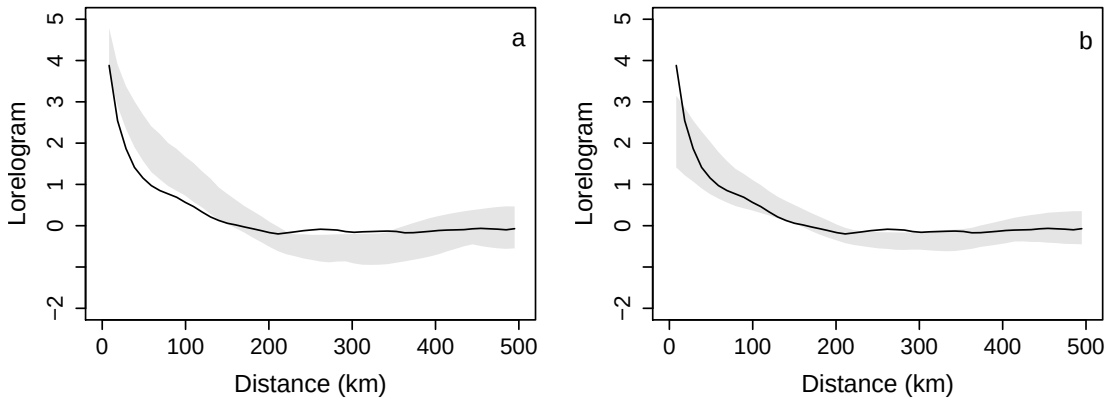

Figure 9.6.: Lorelogram for the precipitation case on June 22nd, 2011 at 16 UTC. The black curve is the radar lorelogram and the grey shaded area represents the lorelograms computed for 50 realisations of occurrence/non-occurrence of precipitation ≥ 0.1 mm/h obtained with a) THR and b) BNS.

The occurrence of precipitation tends to be slightly underestimated by HIRAIN (BIAS < 1). Both methods tend to fail to reproduce the large rainy area in the eastern south of Germany. This behaviour is represented by the realisations c) and e) in Figure 9.5. This is probably due to a discrepancy between radar and station observations, which in contrast to radar did not measure precipitation in this region.

The POD indicates that THR is able to detect 65% of the precipitation occurrences observed by radar, while BNS detects 59% of them. This is, however, obtained on the cost of 23 % false alarm for BNS and THR. In both methods, the odds of making a hit is greater than the odds of making a false alarm. They are also able to detect occurrence of precipitation accounting for hits to happen by chance and to separate between occurrence and non-occurrence of precipitation ≥ 0.1 mm/h (table 9.1).

Figure 9.5 reinforces what is shown by the categorical statistics: THR has better performance in comparison to BNS. The fields generated with THR are more homogeneous and similar to radar than those produced by BNS. The noisier spatial structure obtained from BNS is due to the fact that the method does not account for spatial dependence in the process of occurrence/non-occurrence of precipitation. Opposite to BNS, THR thresholds the fields of $\mathbf{s}^p$, which in turn accounts for

the spatial dependence in the latent Gaussian process and thus, produces more homogeneous fields.

The spatial dependence structure is further investigated by the means of a lorelogram (Figure 9.6). The radar field has a large log odds ratio of about 4 for neighbor locations. This large dependence is well reproduced by THR, whereas BNS largely underestimates nearby spatial dependence. As already suggested visually in Figure 9.5, this shows that ignoring spatial dependence at nearby locations is less appropriate than assuming conditional dependence to generate occurrence/non-occurrence of precipitation.

Opposite to BNS, THR tends to overestimate the dependence at distances between 30 km and 150 km approximately. This is, however, not a general feature of THR, but a particularity of this precipitation case. During other times, the lorelograms for THR largely differ (not shown). The underestimation by BNS of the nearby dependencies, however, is general.

9.2.2. Category B

9.2.2.1. Categorical statistics

The categorical statistics for the precipitation cases in category B show more variability and reach also higher values in comparison to the cases in category A (Figures 9.7 and 9.8).

There are cases in category B, for which the realisations detect less than 30% of the occurrences observed by radar, whereas in category A POD is equal or greater than 30% for all events. The FAR can reach up to 80% in category B and the log odds ratio, ETS and HK showed also negative values.

The precipitation event on January 18th, 2011 at 06 UTC is a case that shows good BIAS, however with poor results for the other categorical statistics. The predicted occurrences of precipitation is very similar to the observations (BIAS of about 1), however they are not necessarily located in the same region, as indicated by a POD of about 0.35 and FAR of 0.65. The ETS, HK and log odds ratio show values very near to zero. Thus, there is poor ability to predict precipitation occurrences, accouting for hits due to chance and to separate between occurrence

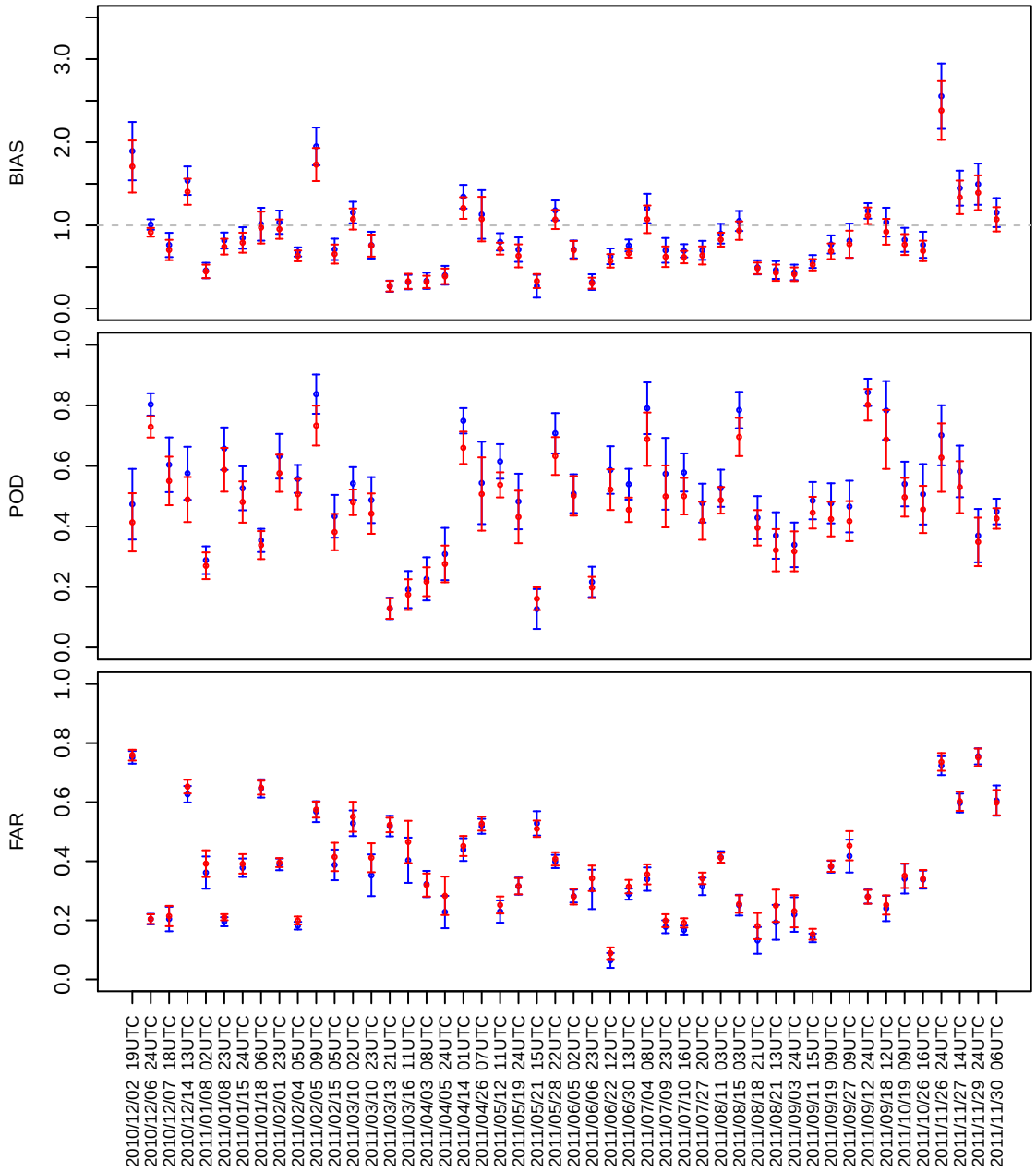

Figure 9.7.: Same as Fig. 9.3 but for the precipitation cases in category B.

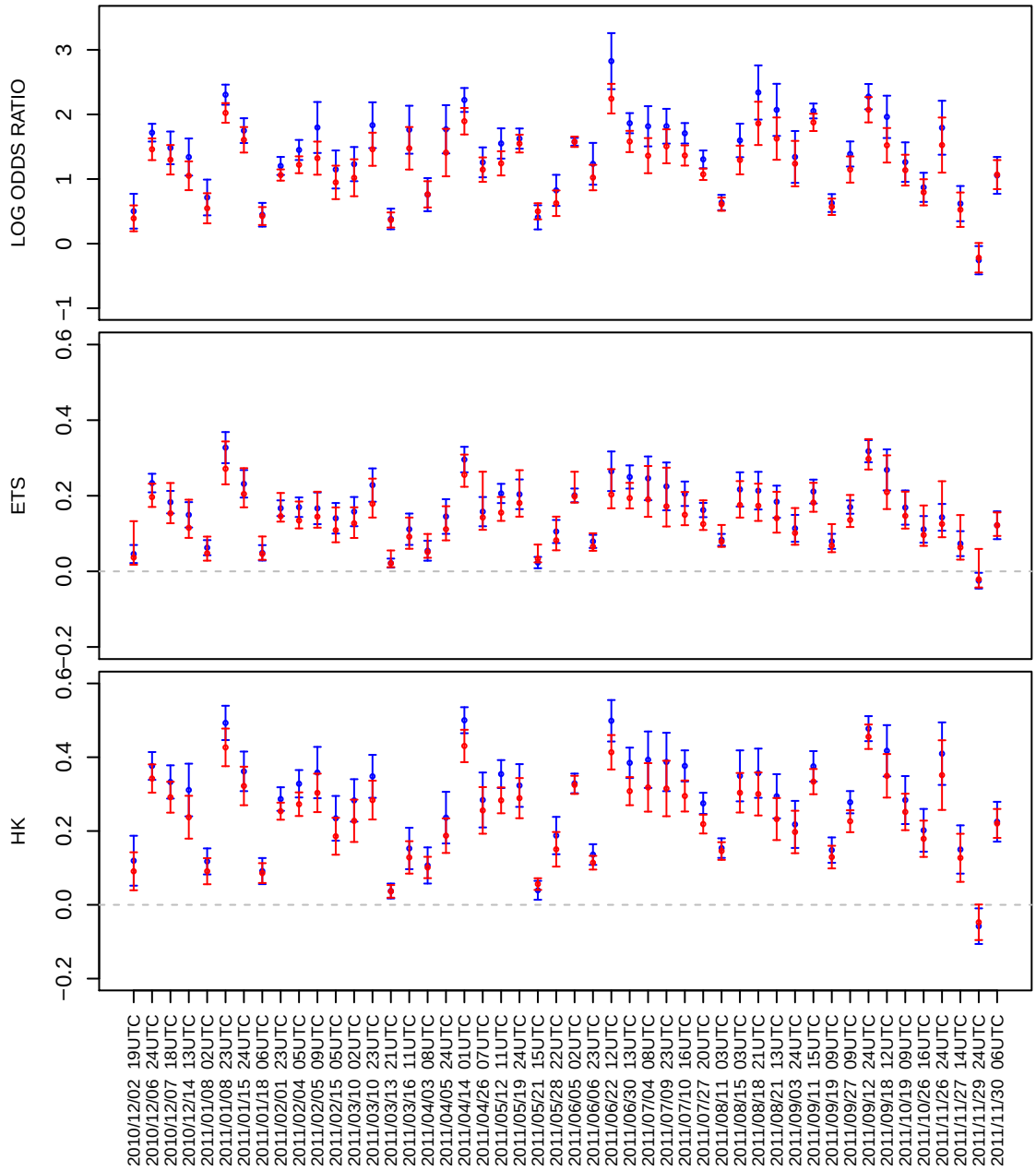

Figure 9.8.: Same as Fig. 9.3 but for the precipitation cases in category B and for the categorical statistics log odds ratio, ETS and HK.

and non-occurrence of precipitation. Besides, the odds of making a hit is slightly greater than the odds of making a false alarm. During this case, 10 synoptic

stations located in western Germany observed precipitation $\geq$ 0.1 mm/h. The covariate is in good agreement with the stations. However, radar observed precipitation in a smaller region (not show). Besides, radar detects precipitation in the southeast part of the country while the stations do not.

Opposite to that is seen on May 21st, 2011 at 15 UTC for which all categorical statistics are poor. BIAS is about 0.3, POD is not larger than 0.2 and FAR is approximately 0.55. There is poor skill regarding the correctly predicted $\mathbf{y}^p$, accounting for hits to happen due to chance (ETS near zero). The ability to separate between occurrence and non-occurrence is also poor (HK near zero). Besides, the low log odds ratio shows that the odds of making a hit is not much higher than the odds of making a false alarm. Radar observed precipitation in a much larger area than the stations (not shown). The covariate is in better agreement with radar than with the stations as well. In this case, the categorical statistics are poor due to poor information given as input to the model.

The precipitation occurrence on November 26th, 2011 at 24 UTC shows the highest BIAS in category B (about 2.5). The POD is approximately 0.6 and 0.75 for BNS and THR, respectively on the costs of a FAR of 0.75. The correct predictions accounting for hits to happen by chance produced ETS of about 0.15. There is good ability to differentiate between occurrence and non-occurrence as showed by the high HK of about 0.4. The log odds ratio shows values of about 1.5.

Radar detects precipitation in a smaller area than the stations (not shown). The covariate does not provide good information to the model. There is not much difference in the IRBT between rainy and non-rainy areas. For this event, MCMC simulations show poor convergence of the chains, what might results in less good predictions.

9.2.2.2. Spatial dependence structure

The spatial dependence structure is investigated in more detail for the precipitation events on July 4th, 2011 at 08 UTC (Figure 9.9) and on December 14th, 2010 at 13 UTC (in appendix D). For the event in July, the rainy area observed by the synoptic stations agrees with the rainy area detected by radar. The correspondent categorical statistics are summarized in table 9.2.

Although the differences in the statistics between THR and BNS are small, THR

Table 9.2.: Categorical statistics for the precipitation case on July 04th, 2011 at 08 UTC.

Statistics	THR	BNS
BIAS	1.2	1.07
POD	0.79	0.69
FAR	0.34	0.36
Log OR	1.82	1.36
ETS	0.25	0.19
HK	0.39	0.32

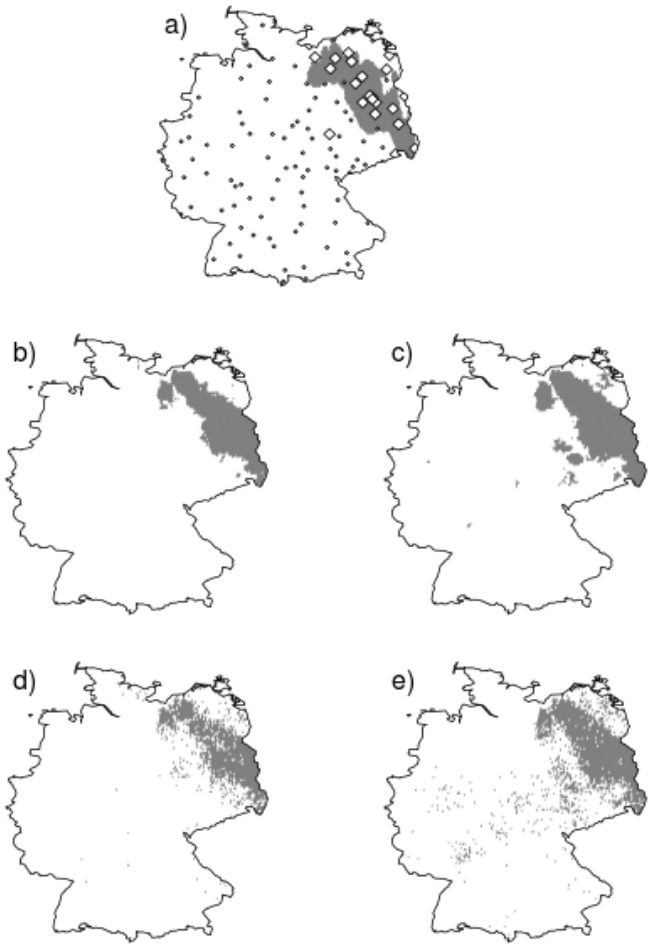

Figure 9.9.: Same as Fig. 9.5 but for the precipitation case on July 04th, 2011 at 08 UTC.

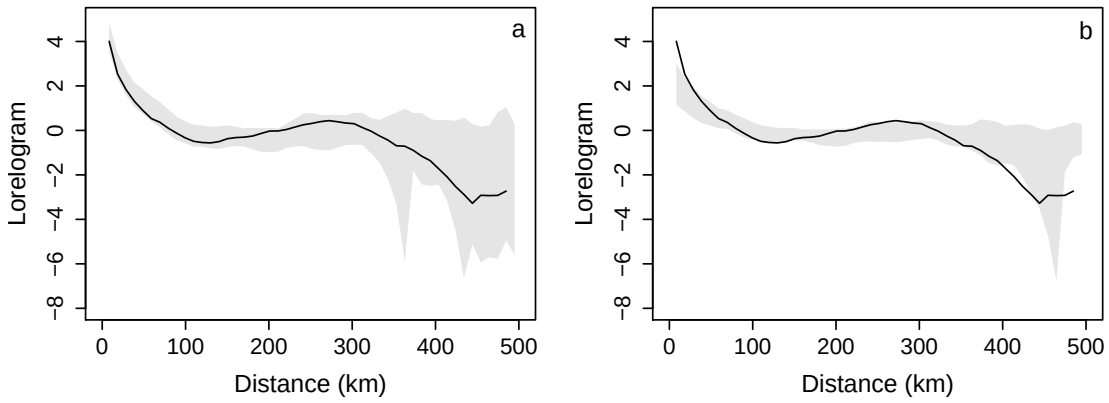

Figure 9.10.: Same as Fig 9.6 but for the precipitation case on July 04th, 2011 at 08 UTC.

presents better values than BNS, except for BIAS. THR predicts 79% of the observed occurrences on the cost of 34% of false alarm, whereas BNS has a lower POD of 69% on the cost of a higher FAR of 36%. This is also notable by comparing the realisations in Figure 9.9 with radar. THR agrees better with radar than BNS.

There is skill in the realisations regarding the correct predictions, accounting for hits to happen by chance. This is given by ETS of 0.25 for THR and and 0.19 for BNS. The ability to separate between occurrence and non-occurrence of precipitation is also best when THR is used (HK of 0.30 for THR and 0.32 for BNS).

The spatial structure observed by radar is in general very well captured by the realisations in both methods (Figure 9.10). THR is able to reproduce the spatial structure at all distances, whereas BNS disagrees with radar at nearby locations and at approximately 420 km. It underestimates the dependency at nearby locations and overestimates them at about 420 km. The latter is, however, a particularity of this precipitation case. For most distances, radar can be considered as one realisation of HIRAIN, regardless of which method is used to generate the fields occurrence/non-occurrence.

9.2.3. Category C

9.2.3.1. Categorical statistics

The ability of HIRAIN to correctly predict precipitation occurrences observed by radar decreased in category C in comparison to categories A and B. POD is not greater than 0.6 here (Figure 9.11). The ETS and HK statistics also presented lower values in category C than in categories A and B (Figure 9.12). ETS ranges from 0 to 0.2, while HK reaches up to 0.4. More precipitation events in category C show log odds ratio, ETS and HK near zero than in categories A and B. Thus, the odds of making a hit is more often comparable to the odds of making a false alarm in category C than in the categories A and B. The correct predictions by chance increase and the ability to differentiate between occurence and non-occurrence decreases in category C compared to the others. FAR is comparable to category B and ranges mostly from 0.2 to 0.7.

The precipitation event on March 09th, 2011 at 20 UTC shows an example, where BIAS is approximately 1, however with poor POD and FAR. The POD is about 0.2 and FAR is approximately 0.9. In this case, the number of grid points with observed and predicted precipitation agrees, however the locations with observed and predicted precipitation is different. The predictions show precipitation in a different region than radar (not shown). This is because the training stations observed precipitation at different locations than radar does. Since ETS and HK are both nearly zero, there is not much skill to predict precipitation occurrence accounting for hits to happen by chance and to separate occurrences from non-occurrences.

On March 13th, 2011 at 22 UTC the predictions are strongly underestimated in comparison to radar. BIAS is approximately 0.3, the POD is about 0.1 and the FAR amounts to 0.4. The log odds ratio is greater than zero, whereas ETS and HK are nearly zero. In this case the strong disagreement between predictions and radar is due to the fact that only 9 training stations observed precipitation $\geq$ 0.1 mm/h in the western south of Germany, while radar detected precipitation for a larger area in the west part of the country (not shown).

The event on September 04th, 2011 at 16 UTC presents good categorical statistics. POD is of about 0.5, FAR approximately 0.1 and the log odds ratio is the highest in the category C (near 2). The ETS and HK amount to 0.2 and 0.4, respec

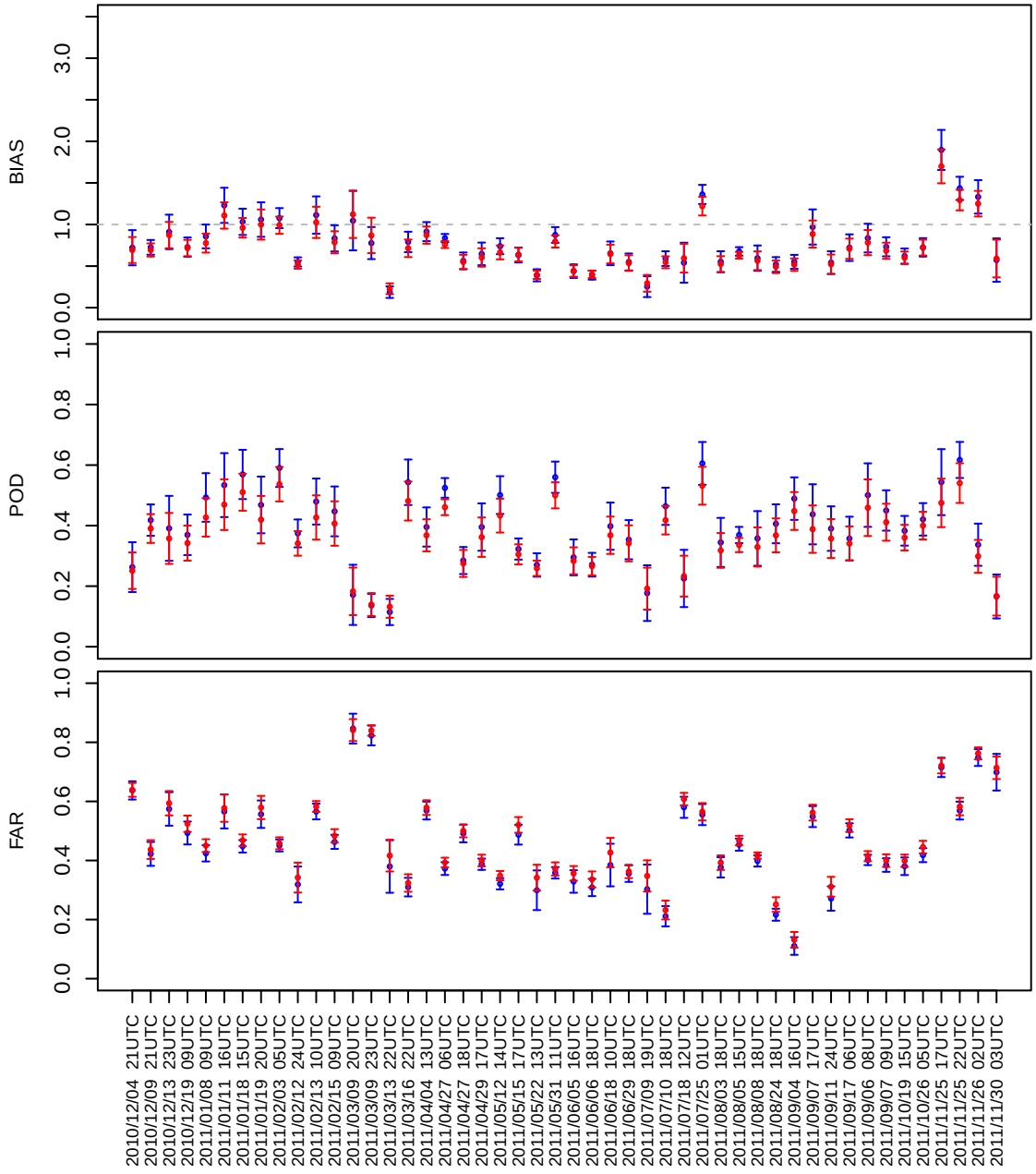

Figure 9.11.: Same as Fig. 9.3 but for the precipitation cases in category C.

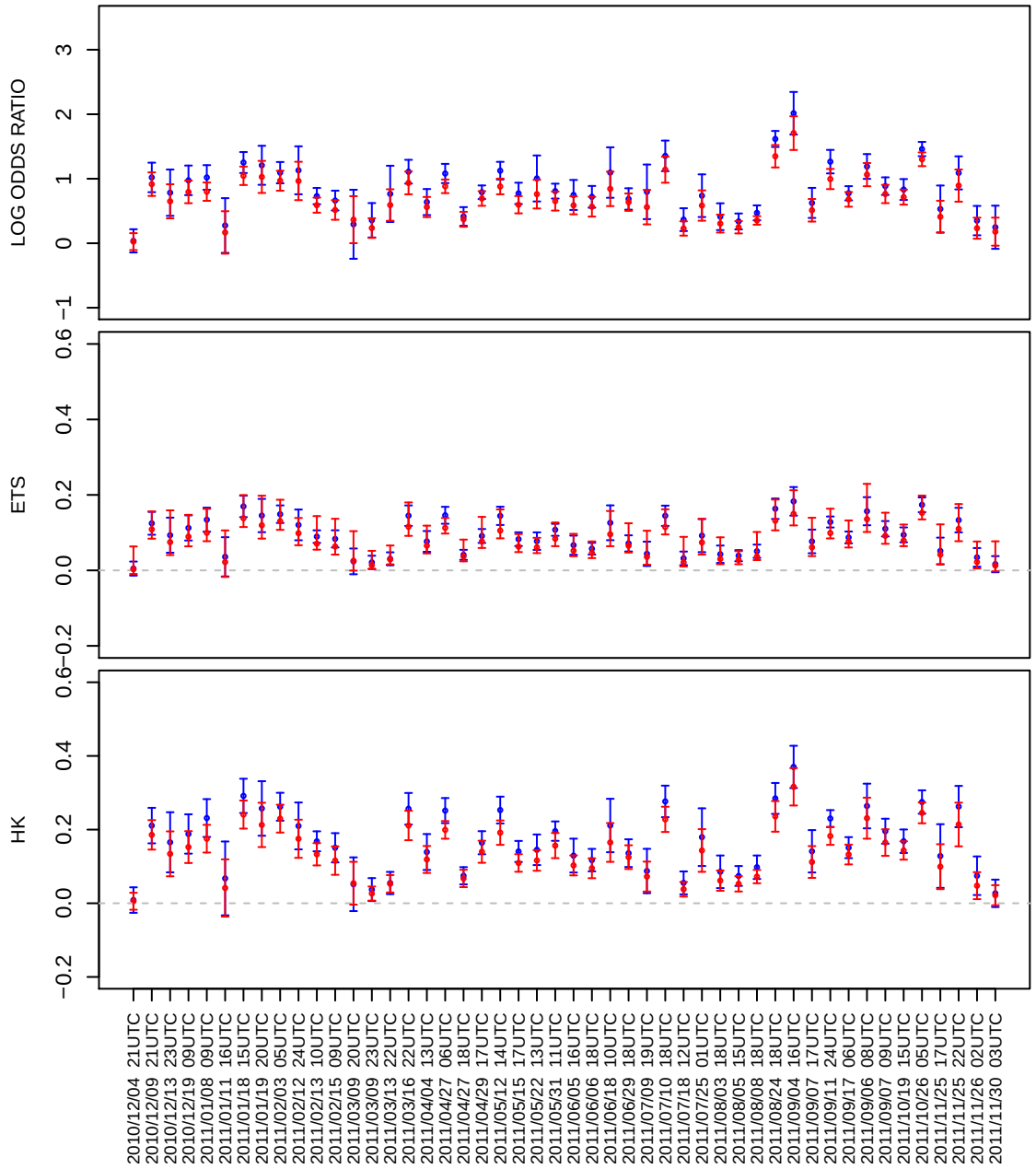

Figure 9.12.: Same as Fig. 9.3 but for the precipitation cases in category C and for the categorical statistics log odds ratio, ETS and HK.

tively. The occurrences of precipitation ≥ 0.1 mm/h is, however, underestimated by HIRAIN (BIAS ~ 0.5).

For this event radar shows precipitation $\geq$ 0.1 mm/h in a large area in the western Germany extending from north to south (not shown). The training stations are in good agreement with radar in the south, whereas the agreement tends to decrease towards the north. Radar presents a larger rainy area there than the stations. The covariate shows stronger agreement with the training stations in the south than in the north area. From these observations, the predicted occurrence/non-occurrence of precipitation $\geq$ 0.1 mm/h shows good agreement with radar in the southern. In the northern the predictions underestimate the radar observations (not shown).

9.2.3.2. Spatial dependence structure

A field of occurrence/non-occurrence of precipitation $\geq$ 0.1 mm/h in category C is shown for October 19th, 2011 at 15 UTC (Figure 9.13) and for June 05th, 2011 at 16 UTC (appendix D). The correspondent categorical statistics for October are summarized in table 9.3.

THR and BNS predict precipitation occurrence $\geq$ 0.1 mm/h in southern and northwestern Germany, however, the locations are displaced in comparison to radar. As shown in chapter 8 there is not significant correlation between precipitation observed by the training stations and the covariate, which might contribute to the disagreement between the simulated fields and radar as well.

The categorical statistics for this case confirm what is seen visually. Precipitation occurrence is underestimated by both methods (BIAS $<$ 1).

THR and BNS predict correctly only 38% and 36%, respectively of the radar observations on the costs of a false alarm of 38% and 40%, respectively. The odds of making a hit is greater than the odds of making a false alarm. However, there is poor ability to predict precipitation occurrence, accouting for hits to happen by chance. The ability to separate the occurrences from the non-occurrences is given by a HK of 0.17 for THR and 0.14 for BNS.

Even though the predicted regions with precipitation $\geq$ 0.1 mm/h do not exactly match the rainy regions observed by radar, the spatial dependency structure of radar is well reproduced by THR and BNS (Figure 9.14). It is emphasized once more the difference between both methods at the nearby locations.

Table 9.3.: Categorical statistics for the precipitation case on October 19th, 2011 at 15 UTC.

Statistics	THR	BNS
BIAS	0.62	0.60
POD	0.38	0.36
FAR	0.38	0.40
Log OR	0.83	0.71
ETS	0.09	0.08
HK	0.17	0.14

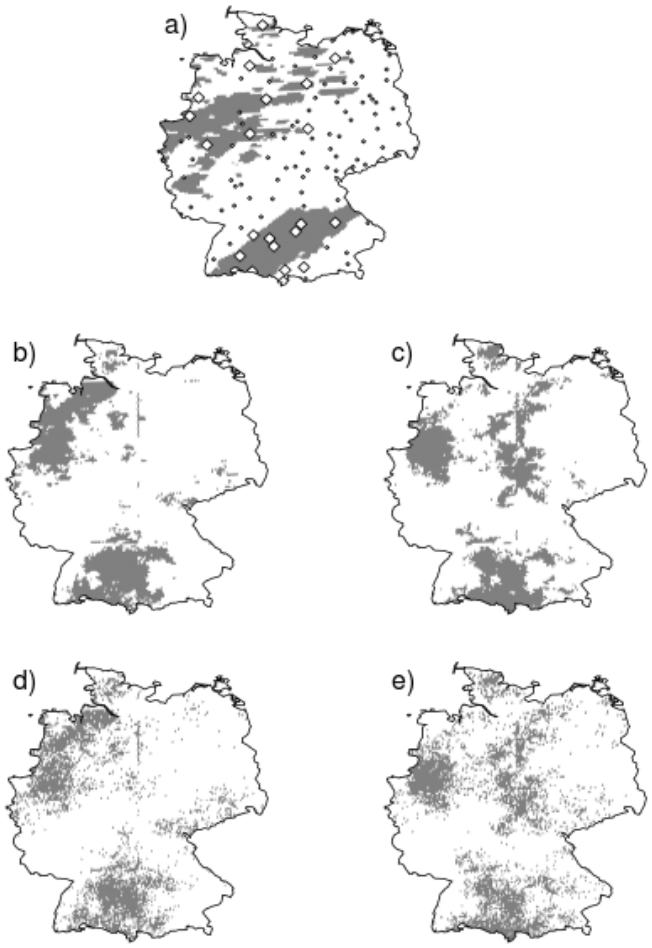

Figure 9.13.: Same as Fig. 9.5 but for the precipitation case on October 19th, 2011 at 15 UTC.

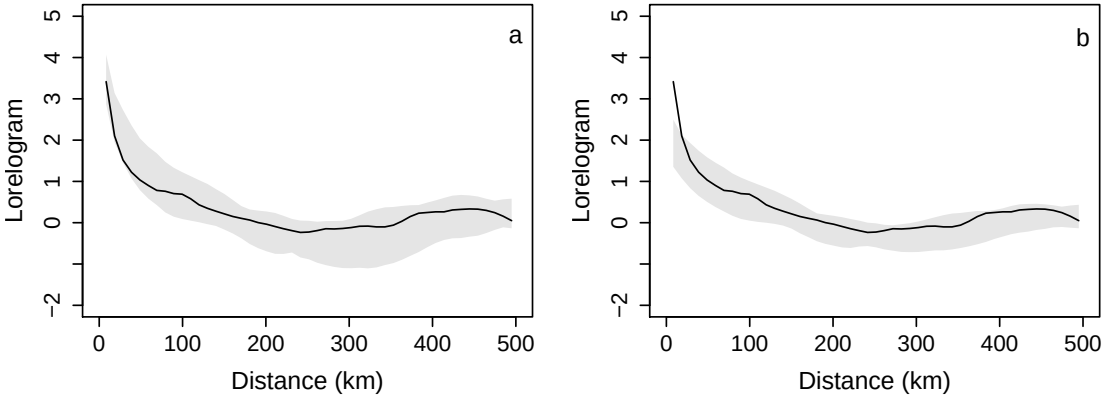

Figure 9.14.: Same as Fig. 9.6 but for the precipitation case on October 19th, 2011 at 15 UTC.

10. HIRAIN extended to multiple thresholds

HIRAIN is now extended to thresholds of precipitation amounts of 0.5 mm/h, 1.0 mm/h, 2.5 mm/h, 5.0 mm/h, 7.0 mm/h and 10.0 mm/h.

HIRAIN starts with the threshold of 0.1 mm/h and simulates $\boldsymbol{s}^p$ and $\pi(\boldsymbol{r}^p)$ (step 1) as well as realisations of $\boldsymbol{y}^p$ (step 2) of the 0.1 mm/h threshold process. The procedure is then repeated for the next lowest threshold, namely 0.5 mm/h. The estimation of the posterior distribution is independent of the estimates of the lower threshold. The observations $\boldsymbol{y}^o$ are now binary values obtained by assigning 1 when rainfall amount is $\geq$ 0.5 mm/h and 0 otherwise. The prior distributions for $\boldsymbol{\beta}$, σ^2 and ϕ are the same as used for the threshold, thr=0.1 mm/h and discussed in Appendix A.

Occurrence/non-occurrence of precipitation above 0.5 mm/h is simulated only at the grid points where there is precipitation $\geq$ 0.1 mm/h. Threshold exceedance of 0.5 mm/h may only occur where the threshold of 0.1 mm/h is exceeded. This assures that the fields are consistent throughout the thresholds, although they are generated independently. This procedure is further repeated until HIRAIN produces fields $\boldsymbol{y}^p$ of every threshold selected in this study.

Since the greatest differences in the ability of HIRAIN to predict $\boldsymbol{y}^p$ in terms of the BSS is among the precipitation thresholds of 0.1 mm/h, 2.5 mm/h and 7.0 mm/h, the evaluation of HIRAIN is presented here only for precipitation $\geq$ 2.5 mm/h and $\geq$ 7.0 mm/h.

The amount of 2.5 mm/h corresponds to moderate rain and occurs less often than the events with lower amounts of precipitation (Figure 8.2). From the thresholds selected in this study, 2.5 mm/h is also the greatest threshold for which precipitation occurrences can still be found in most months during the year.

The occurrences of precipitation $\geq$ 7.0 mm/h are mostly concentrated on the summer months. The spatial distribution of the synoptic stations with precipitation also becomes reduced. Beside that, when this scenario meets the condition,

where the rainy stations are wide distributed over the country, the performance of HIRAIN tends to decrease. These are the characteristics of category C presented in chapter 8. Thus, it is expected that the higher the threshold of precipitation, the more difficult it is to generate $\boldsymbol{y}^p$.

10.1. Probability of precipitation

10.1.1. Model evaluation

The performance of HIRAIN in predicting probability of precipitation $\geq$ 2.5 mm/h is validated against the validation synoptic stations (Figure 3.2 (right)) by means of the BSS.

The median of the BSS in Figure 10.1 amounts to 30%. The BSS slightly increases during the months from July to September of 2011. HIRAIN performs best during July, for which the median of the BSS is about 30%. The BSS of August and September are very similar. From January to June and during October the median of the skill score amounts to about 20%. In December 2010 only 11 hours show precipitation $\geq$ 2.5 mm/h, and the BSS is close to zero, i.e., there is no increase in the skill with respect to climatology. At these 11 hours, 4 to 9 synoptic stations observed precipitation. These stations are located in one region in Germany (as in category B). Besides, they do not present good agreement with the IRBT. At some hours, the stations are concentrated in the border of the country. As seen in chapter 8, all these conditions showed to be very unfavorable for good performance of the BSGLM and explains why the model presented no skill during December.

Occurrence of precipitation $\geq$ 7.0 mm/h is only seen from May to September of 2011. The median of the BSS during these months varies from 0.1 to 0.25. Although the skill scores are low, they indicate skill in the predictions with respect to the climatology. The highest median BSS of 0.25 is seen in July. However, there is also more dispersion of the scores around the median during July than during the other months.

As the precipitation threshold increases, the number of stations with precipitation decreases. As seen in chapter 8, the skill is reduced when the number of observations is below 25, specially when they are wide distributed over the coun-

try.

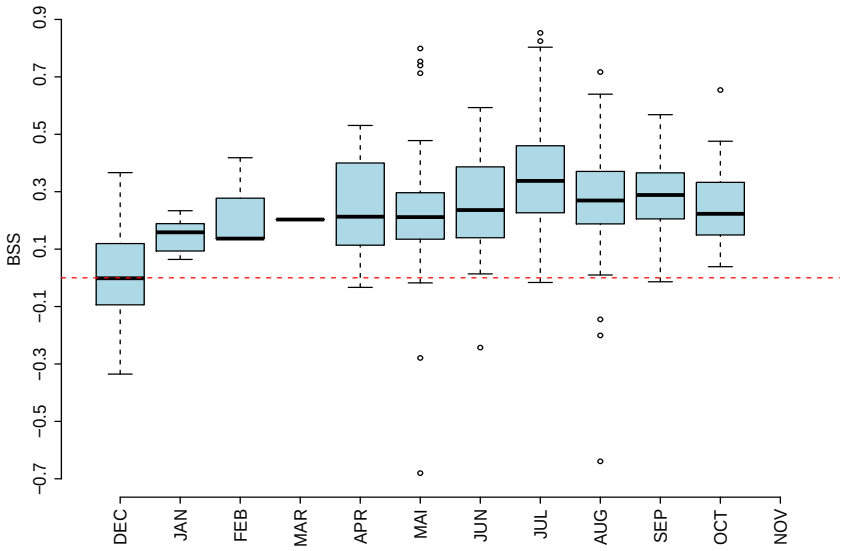

Figure 10.1.: Same as Fig. 8.1 but for probability of precipitation $\geq$ 2.5mm/h.

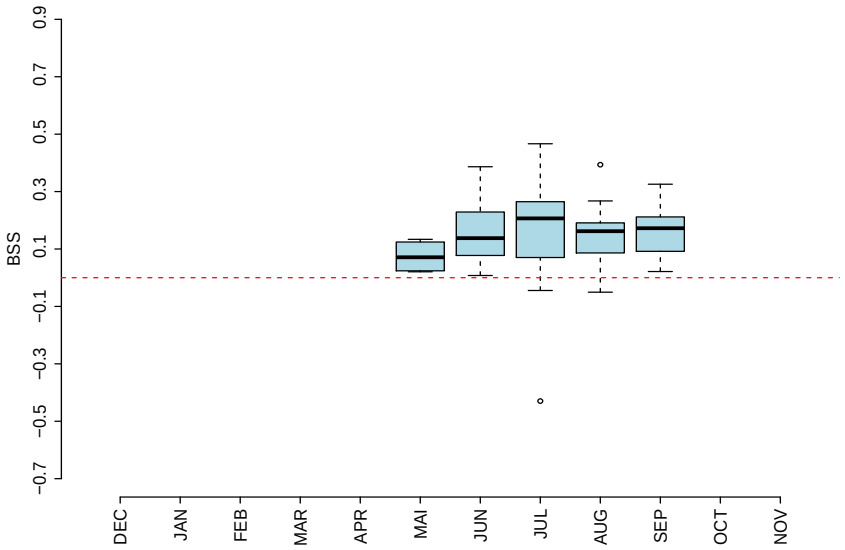

Figure 10.2.: Same as Fig. 8.1 but for probability of precipitation $\geq$ 7.0mm/h.

For the 7.0 mm/h threshold there might not be enough training synoptic stations to represent the rainy area. This provides poorer predictions. Opposite to the

training stations, the network of the validation stations is much denser (Figure 3.2) and it is able to better represent the rainy regions. This might favour the poor BSS for the 7.0 mm/h threshold.

Increasing the number of training stations might contribute to better predictions, since precipitation observed with a denser network would provide enough input data to the BSGLM for higher amounts of precipitation as well. However, increasing the number of training stations already showed to slow down the running time of HIRAIN.

10.1.2. Cases

The simulated probability fields for the thresholds of 2.5 mm/h and 7.0 mm/h are shown here for the cases on June 22nd, 2011 at 16 UTC (category A) and on July 4th, 2011 at 08 UTC (category B).

The BSS amounts to 55% and 72%, respectively. There is a slight decrease in the ability of the BSGLM to simulate $\pi(\boldsymbol{r}^p)$ with the increase of the precipitation threshold. For precipitation amount $\geq$ 0.1 mm/h the BSS amounts to 60% and 73%, respectively (tables 8.1 and 8.2).

In order to compute the correlation coefficient between precipitation and IRBT, precipitation amount $<$2.5 mm/h is replaced here by zero. The amounts $\geq$ 2.5 mm/h are retained and the correlation is calculated using this new precipitation dataset. The same logic is followed with the threshold of 7.0 mm/h. The idea behind this is to assess whether there is a relationship between precipitation occurrrence/non-occurence $\geq$ 2.5 mm/h and $\geq$ 7.0 mm/h and IRBT. The summary of the significance test can be found in Appendix D.

For the case on June 22nd, 2011 at 16 UTC, there is a significant correlation coefficient of -0.29 between precipitation and IRBT. The IRBT at the rainy stations presents mean equal 224.5 K, whereas for the non-rainy observations the mean IRBT is 235.3 K. The model agrees well with the validation stations in the region going from the central towards the northern Germany and in part of the south-western region. There is precipitation $\geq$ 2.5 mm/h and low IRBT in these areas (Figure 10.3(top left)). The realisations show probability of precipitation $\pi(\boldsymbol{r}^p) \geq$ 0.8 there. Between both areas, in the central part of the country, the simulated probability is $\pi(\boldsymbol{r}^p) \geq 0.8$, however, the validation stations do not show precipita-

Table 10.1.: Selected cases with precipitation $\geq$ 2.5 mm/h and the corresponding BSS.

Cases	BSS
22. June 2011, 16 UTC	0.55
4. July 2011, 08 UTC	0.72

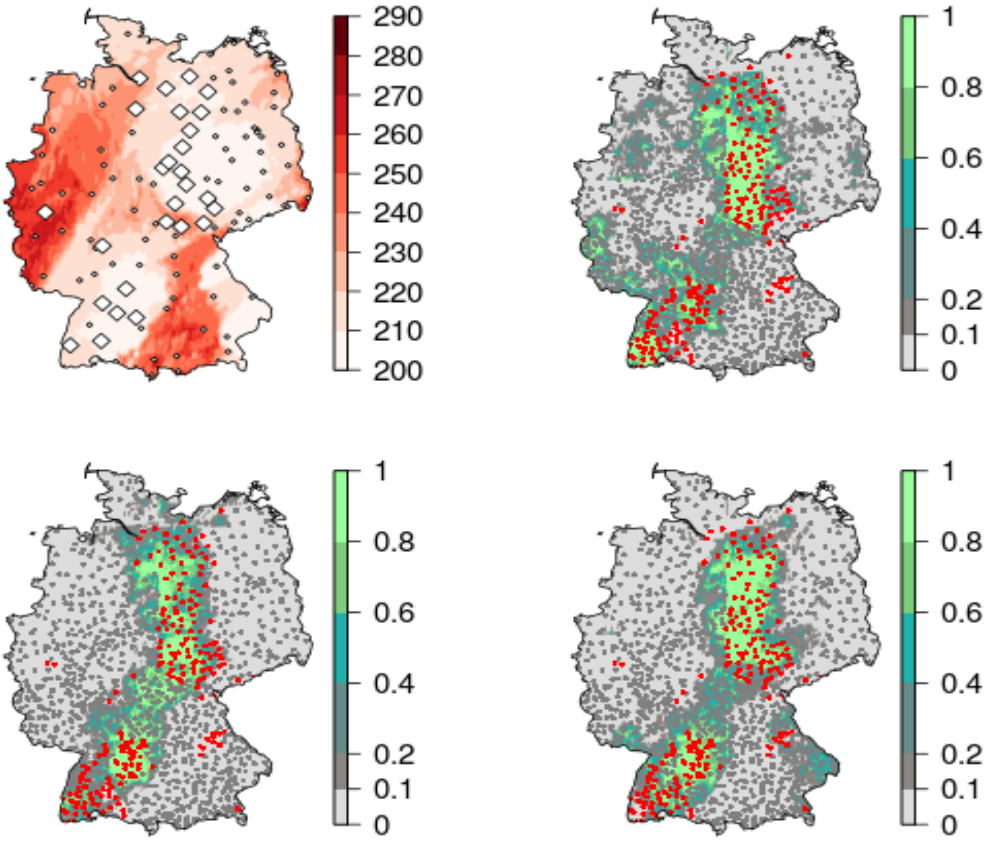

Figure 10.3.: Same as Fig. 8.7 but for precipitation $\geq$ 2.5 mm/h on June 22nd, 2011 at 16 UTC.

tion. IRBT and the training stations do not agree very well there (Figure 10.3 (top left)). While IRBT presents very low values, the observations show occurrence and non-occurrence of precipitation $\geq$ 2.5 mm/h in the region.

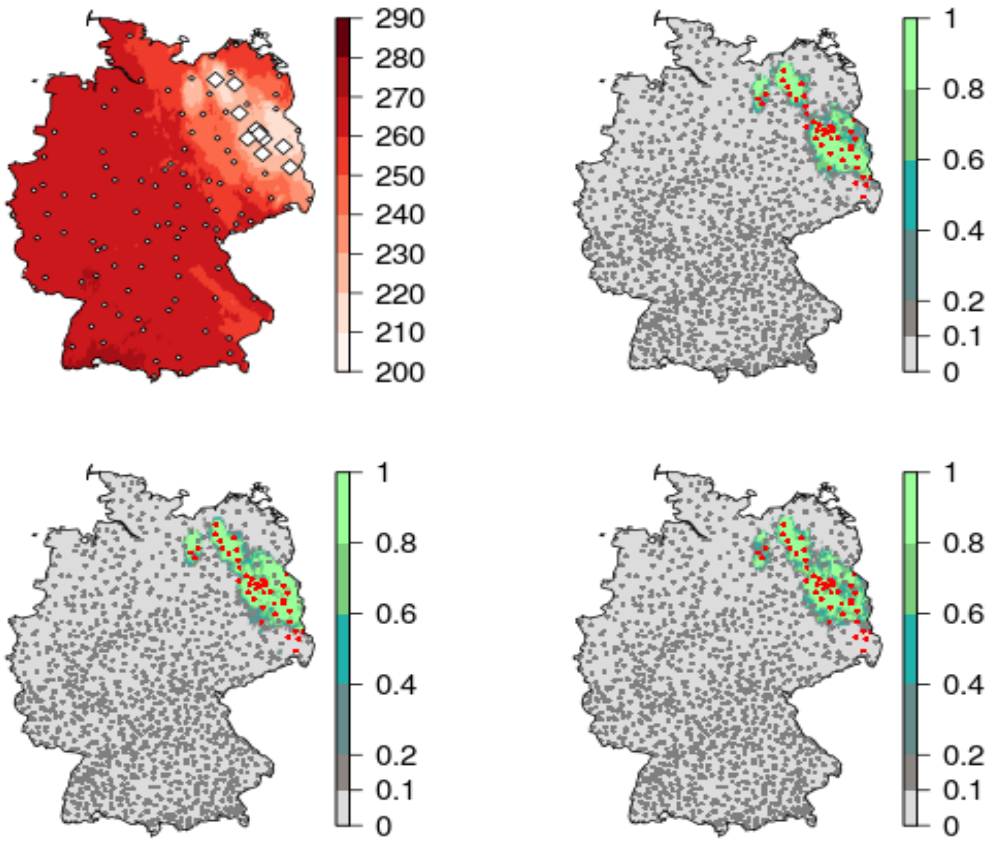

Figure 10.4.: Same as Fig. 8.7 but for precipitation $\geq$ 2.5 mm/h on July 4th, 2011 at 08 UTC.

Opposite feature is seen in the southeastern Germany, where there is precipitation detected by the validation stations, however the simulations show very low probability of occurrence ($\pi(\boldsymbol{r}^p) \leq 0.1$ for most of the realisations). This was already observed for precipitation $\geq$ 0.1 mm/h and is also found for the other precipitation thresholds between 0.1 mm/h and 2.5 mm/h (not shown).

For the precipitation event of category B on July 4th, 2011 at 08 UTC there is a high correlation coefficient between precipitation and IRBT of -0.65. The difference between the covariate at rainy and non-rainy station locations is notable. Mean IRBT of 235 K is found for rainy locations, whereas for non-rainy stations the mean is 275.1 K.

The simulations predicted probability $\pi(\boldsymbol{r}^p) \geq 0.8$ at locations where the validation stations detected precipitation $\geq$ 2.5 mm/h, except at 4 locations near

Table 10.2.: Selected cases with precipitation $\geq$ 7.0 mm/h and the corresponding BSS.

Cases	BSS
22. June 2011, 16 UTC	0.28
4. July 2011, 08 UTC	0.47

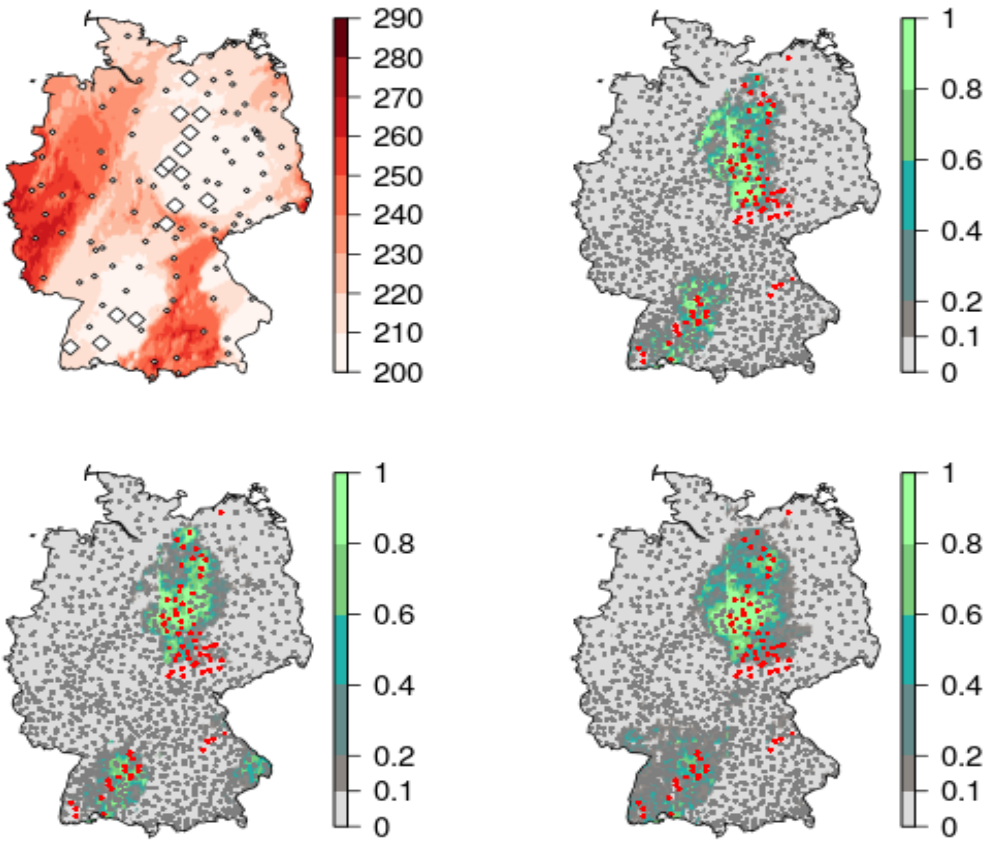

Figure 10.5.: Same as Fig. 8.7 but for precipitation $\geq$ 7.0 mm/h on June 22nd, 2011 at 16 UTC.

the eastern border, where stations detected precipitation and the probability is predicted to be $\pi(\boldsymbol{r}^p) \leq 0.1$ (Figure 10.4). Validation stations with no precipitation are located in areas with probability $\pi(\boldsymbol{r}^p) \leq 0.1$.

The probabilities of precipitation $\geq$ 7.0 mm/h show BSS of 0.28 and 0.47 for the events on June 22nd, 2011 at 16 UTC and on July 04th, 2011 at 08UTC,

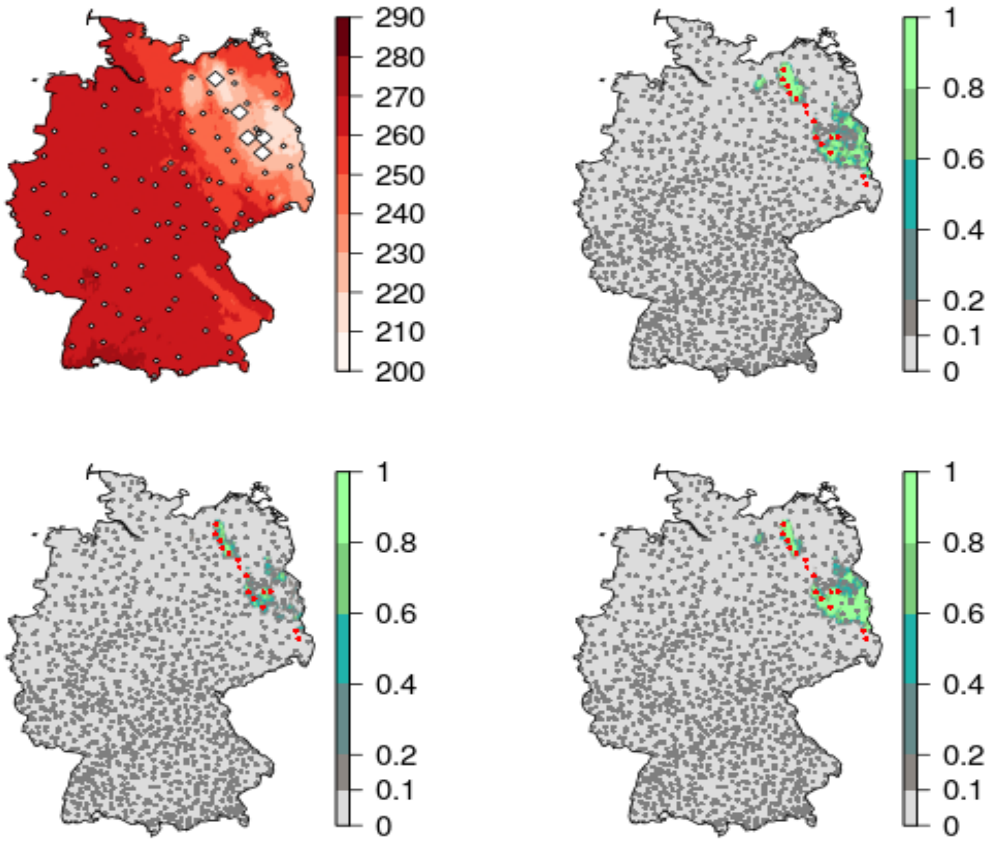

Figure 10.6.: Same as Fig. 8.7 but for precipitation $\geq$ 7.0 mm/h on July 4th, 2011 at 08 UTC.

respectively (Table 10.2). In comparison to lower thresholds, the performance of the BSGLM decreased considerably.

On June 22nd, 2011 at 16 UTC there is a significant correlation coefficient of -0.27 between precipitation and IRBT. The IRBT averaged in space is 219.1 K (234.8 K) for the rainy (non-rainy) locations.

The predictions obtained with the BSGLM show probabilities $\pi(\boldsymbol{r}^p) \geq 0.8$ mostly for the central-north part of Germany and in the southwest, where precipitation $\geq$ 7.0 mm/h is also observed by the validation stations (Figure 10.5). Opposite to that, in the southeast and in the central part, the predicted probability is $\pi(\boldsymbol{r}^p) \leq$ 0.1, but the validation stations observed precipitation.

On July 04th, 2011 at 08 UTC the correlation coefficient between precipitation and IRBT is significant and equal -0.49. The spatial average of the IRBT is 232.8 K for rainy and 273.2 K for non-rainy locations.

Most validation stations with precipitation $\geq$ 7.0 mm/h are located in areas with predicted probability $\pi(\boldsymbol{r}^p) \geq 0.8$ (Figure 10.6). However, there are in the eastern border validation stations with precipitation where the probability ≤ 0.1, and stations with precipitation $<$ 7.0 mm/h in areas with probability ≥ 0.8.

As seen in chapter 8, it is challenging to estimate correctly the parameter ϕ when the number of observations with precipitation is reduced. Beside that, in the area with very low IRBT there are also training stations with no precipitation (Figure 10.6 (top left)). This case reinforces the fact that the number of rainy stations and the agreement between precipitation and covariate play important roles in the model's performance.

10.2. Occurrence/non-occurrence of precipitation

The predicted fields of $\mathbf{y}^p$ for precipitation $\geq$ 2.5 mm/h and $\geq$ 7.0 mm/h are discussed here. As shown for the precipitation threshold of 0.1 mm/h, the probability threshold P_t for THR is defined first. Occurrence/non-occurrence of precipitation $\geq$ 2.5 mm/h and $\geq$ 7.0 mm/h is generated using THR and BNS. This section closes with the validation of $\mathbf{y}^p$ against meteorological radar.

10.2.1. Probability threshold P_t

The probability threshold P_t is obtained here as described in the section 9.1.

The EBS for precipitation $\geq$ 2.5 mm/h (Figure 10.7) is lower than the EBS for the previous thresholds, namely 0.1 mm/h, 0.5 mm/h and 1.0 mm/h (not shown). As the threshold increases the number of non-occurrences also increases. The large number of correctly predicted non-occurrences contributes to the decrease of the EBS. The median of the EBS is the highest, with the value of 0.05 when the probability of 0.1 is used to separate occurrence from non-occurrence of precipitation $\geq$ 2.5 mm/h. The median decreases to approximately 0.025 as the probability increases from 0.1 to 0.4. Probabilities greater than 0.4 show approximately a constant median of EBS of about 0.025.

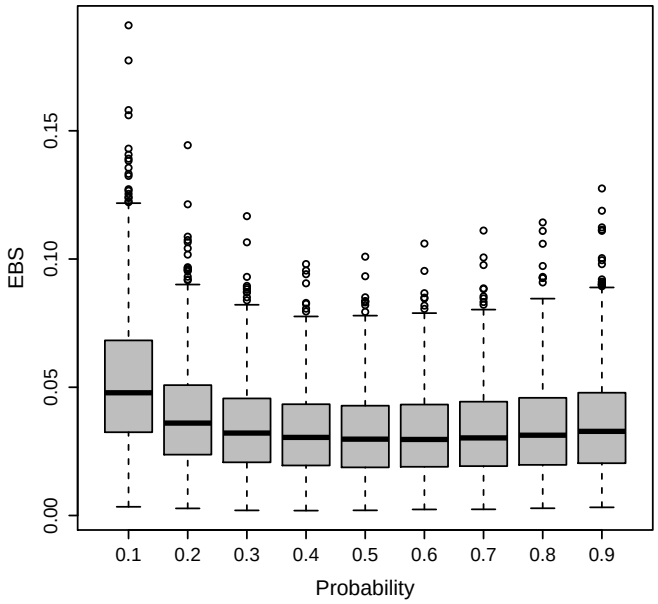

Figure 10.7.: Same as Fig. 9.1 but for precipitation $\geq$ 2.5mm/h.

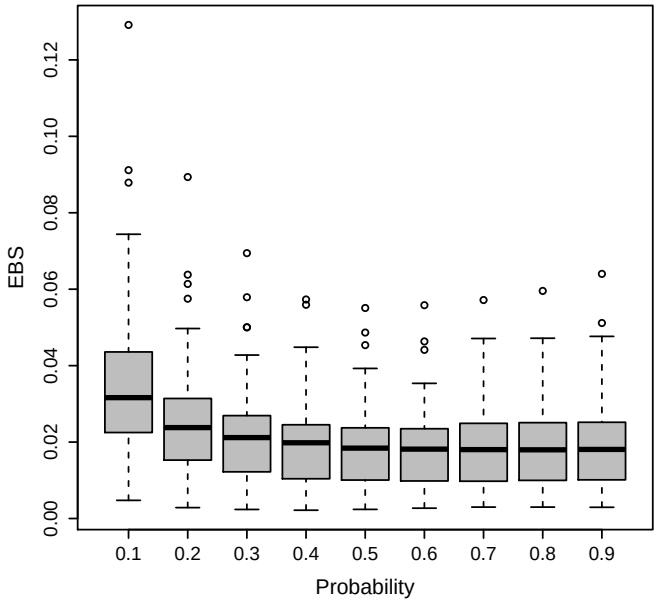

Figure 10.8.: Same as Fig. 9.1 but for precipitation $\geq$ 7.0mm/h.

For precipitation $\geq$ 7.0 mm/h EBS shows even lower values (Figure 10.8) than the EBS for the previous thresholds (2.5 mm/h and 5.0 mm/h). The median of the EBS decreases from 0.03 to 0.02 as the probability increases from 0.1 to 0.4 and remains constant for probabilities $\geq$ 0.4.

The categorical statistics are presented in the Figures 10.9 and 10.10. The median of the log odds ratio ranges between 2 and 3 for both thresholds. Although there is not significant difference, there is a slight increase in the ratio as the probabilitiy increases. This is, however, more evident for precipitation amount $\geq$ 7.0 mm/h. For this case, the uncertainity in the log odds ratio and the number of outliers also decreases as the probability approaches 0.9. Hence, as for the lower thresholds of precipitation, the odds of making a hit is greater than the odds of making a false alarm for all probability thresholds.

The other statistics, namely POD, FAR and BIAS change significantly as the probability used as threshold to define occurrence/non-occurrence of precipitation increases.

In general precipitation $\geq$ 2.5 mm/h (Figures 10.9 (top right)) shows lower POD than the previous thresholds (for instance precipitation $\geq$ 0.1 mm/h in Figure 9.2). POD for precipitation $\geq$ 7.0 mm/h (Figure 10.10 (top right)) is also lower than for the previous thresholds. This statistics decreases faster with the increase of the probability.

The FAR increases as the threshold of precipitation increases (Figures 9.2, 10.9 and 10.10 (bottom left)). While the median of FAR ranges between 0.3 to 0.6 for threshold of 0.1 mm/h, this statistics varies from 0.5 to 0.8 for precipitation $\geq$ 2.5 mm/h and varies around 0.8 for precipitation $\geq$ 7.0 mm/h.

BIAS shows that the number of forecast events is greater than the observed events when the probabilities of 0.1 and 0.2 are used to define occurrence/non-occurrence of precipitation $\geq$ 2.5 mm/h and $\geq$ 7.0 mm/h (Figures 10.9 and 10.10 (bottom right)). The opposite is seen for probabilities $\geq$ 0.5, i.e., the number of forecast events is lower than the number of observed events. For both thresholds BIAS decreases as the P_t increases. The best BIAS is obtained when probabilities between 0.3 and 0.4 are used to define occurrence/non-occurrence of precipitation.

From the results provided by EBS and the categorical statistics the following is concluded. BIAS presents the best values when the probability is between 0.3 and

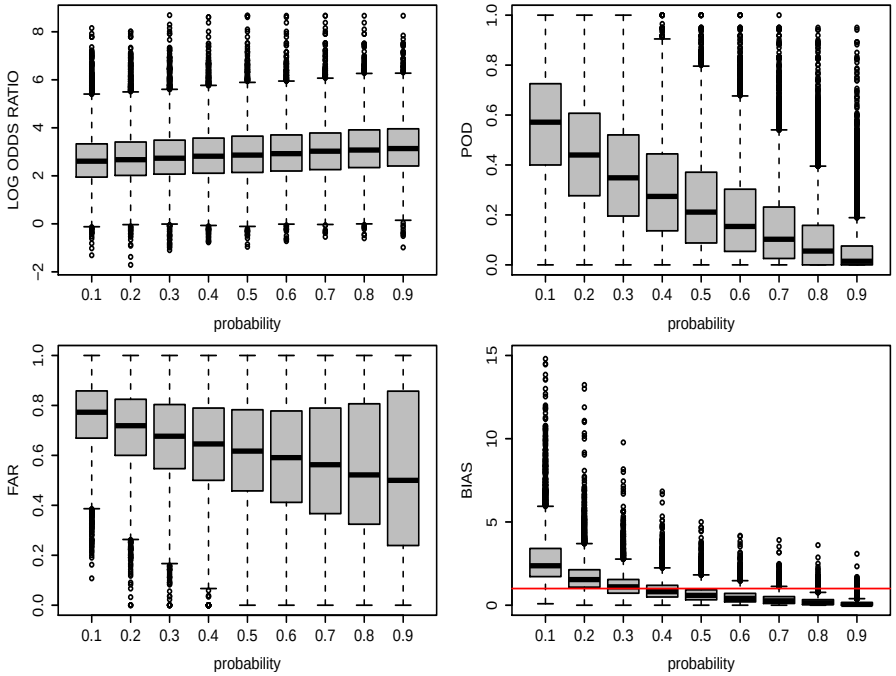

Figure 10.9.: Same as Fig. 9.2 but for precipitation $\geq$ 2.5mm/h.

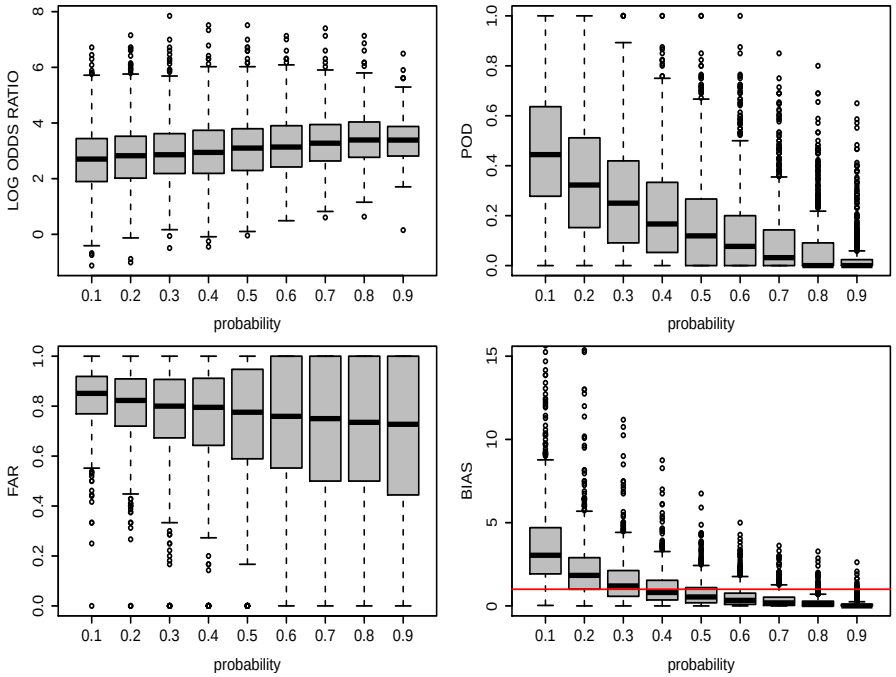

Figure 10.10.: Same as Fig. 9.2 but for precipitation $\geq$ 7.0mm/h.

0.4. Comparing POD and FAR, the appropriate probability should present the highest possible value of POD with the lowest possible value of FAR. The box-and-whiskers plots show that none of the extreme probabilities satisfy this requirement, however, that probabilities between 0.2 and 0.4 are the most appropriate. The log odds ratio increases as the probability threshold increases, although there is no significant differences. Based on this evaluation, the probability of 0.4 is the one that best defines the occurrence/non-occurrence of precipitation $\geq$ 2.5 mm/h and $\geq$ 7.0 mm/h. As for the lower thresholds of precipitation amount, this probability is further used as P_t in THR.

10.2.2. Validation against meteorological radar

10.2.2.1. Cases

Occurrence/non-occurrence of precipitation $\geq$ 2.5 mm/h and $\geq$ 7.0 mm/h are presented here for June 22nd, 2011 at 16 UTC and for July 4th, 2011 at 08 UTC (Figures 10.11, 10.13, 10.15 and 10.17).

There is good agreement between the fields of $\boldsymbol{y}^p$ generated with THR and BNS and radar for precipitation $\geq$ 2.5 mm/h in the event in June (Figure 10.11). This is confirmed by the categorical statistics summarized in table 10.3. BIAS shows that the number of predicted precipitation occurrences is similar to the observed number in both methods (BIAS of 1.07 for THR and 1.02 for BNS). The performance of HIRAIN is for this precipitation threshold better than for precipitation $\geq$ 0.1 mm/h. The latter presented BIAS of 0.85 for THR and 0.77 for BNS.

In terms of the correct predictions HIRAIN performs less good for precipitation $\geq$ 2.5 mm/h than for lower thresholds. THR detects 55% and BNS 47% of the observed events of precipitation $\geq$ 2.5 mm/h on the cost of a false alarm of 48% for THR and 53% for BNS. However, the categorical statistics log odds ratio, ETS and HK are better for precipitation $\geq$ 2.5 mm/h than for precipitation $\geq$ 0.1 mm/h.

The ability to detect occurrence of precipitation accounting for hits to happen by chance produced an ETS of 0.25 for THR and 0.19 for BNS, while the ability to separate between occurrence and non-occurrence of precipitation $\geq$ 2.5 mm/h is indicated by a HK of 0.41 and 0.32, respectively.

THR has better performance in reproducing the spatial dependence structure ob-

Table 10.3.: Categorical statistics for precipitation $\geq$ 2.5 mm/h on June 22nd, 2011 at 16 UTC.

Statistics	THR	BNS
BIAS	1.07	1.02
POD	0.55	0.47
FAR	0.48	0.53
Log OR	2.01	1.62
ETS	0.25	0.19
HK	0.41	0.32

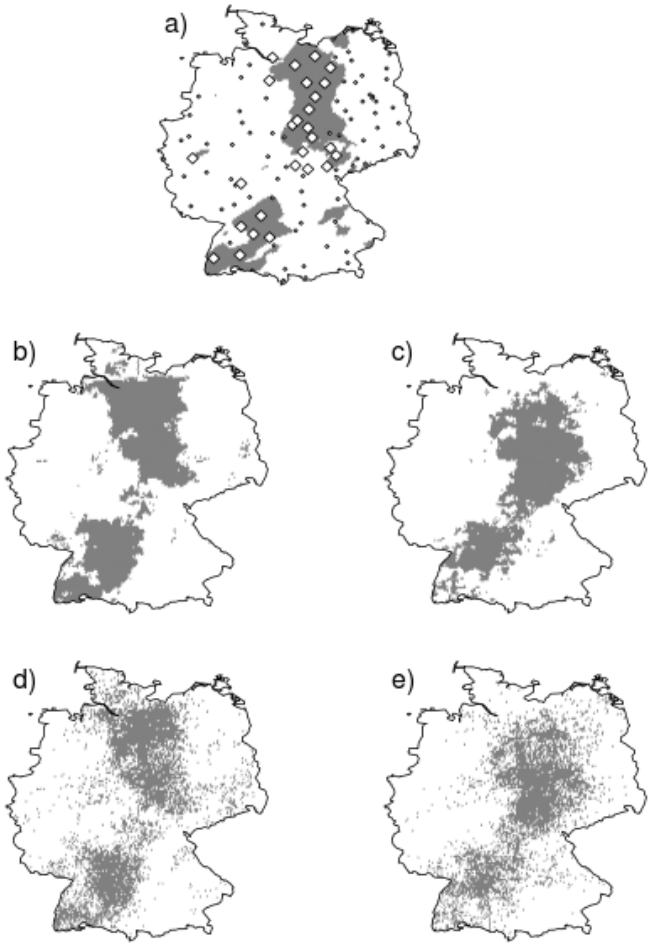

Figure 10.11.: Same as Fig. 9.5 but for occurrence/non-occurrence of precipitation $\geq$ 2.5 mm/h.

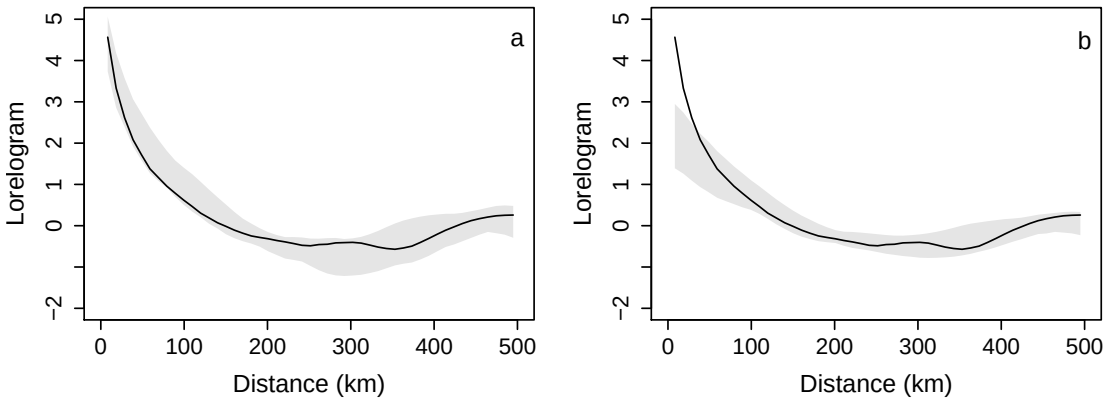

Figure 10.12.: Same as Fig. 9.6 but for the precipitation $\geq$ 2.5 mm/h.

served by radar than BNS (Figure 10.12). Similar to the 0.1 mm/h threshold, the radar field has here a large log odds ratio of about 4.5 for neighbor locations, which is well reproduced by THR, but underestimated by BNS. Except for the nearby distances, both method reproduced very well the spatial structure of radar.

For the precipitation case on July 4th, 2011 at 08 UTC the area with precipitation $\geq$ 2.5 mm/h observed by radar is overestimated by THR and BNS (Figure 10.13). This is reinforced by the BIAS $>$ 1 (1.76 for THR and 1.59 for BNS) (table 10.4).

The performance of HIRAIN to reproduce occurrence/non-occurrence of precipitation $\geq$ 2.5 mm/h decreased in comparison to the 0.1 mm/h threshold. The POD indicates that 68% (THR) and 59% (BNS) of the observed events exceeding 2.5 mm/h are correctly predicted on the costs of a FAR of 61% and 62%, respectively. Comparing these statistics to those for precipitation $\geq$ 0.1 mm/h, the latter presented better performance. However, the ratio of the odds of making a hit to the odds of making a false alarm is greater for precipitation $\geq$ 2.5 mm/h than for precipitation $\geq$ 0.1 mm/h. The ability to separate between occurrence and non-occurrence of precipitation is also better for the current threshold than for 0.1 mm/h. Predictions for precipitation $\geq$ 2.5 mm/h present an HK of 0.47 (THR) and 0.40 (BNS), whereas for precipitation $\geq$ 0.1 mm/h this statistic is 0.39 and 0.32, respectively.

Table 10.4.: Categorical statistics for precipitation $\geq$ 2.5 mm/h on July 4th, 2011 at 08 UTC.

Statistics	THR	BNS
BIAS	1.76	1.59
POD	0.68	0.59
FAR	0.61	0.62
Log OR	2.12	1.83
ETS	0.22	0.19
HK	0.47	0.40

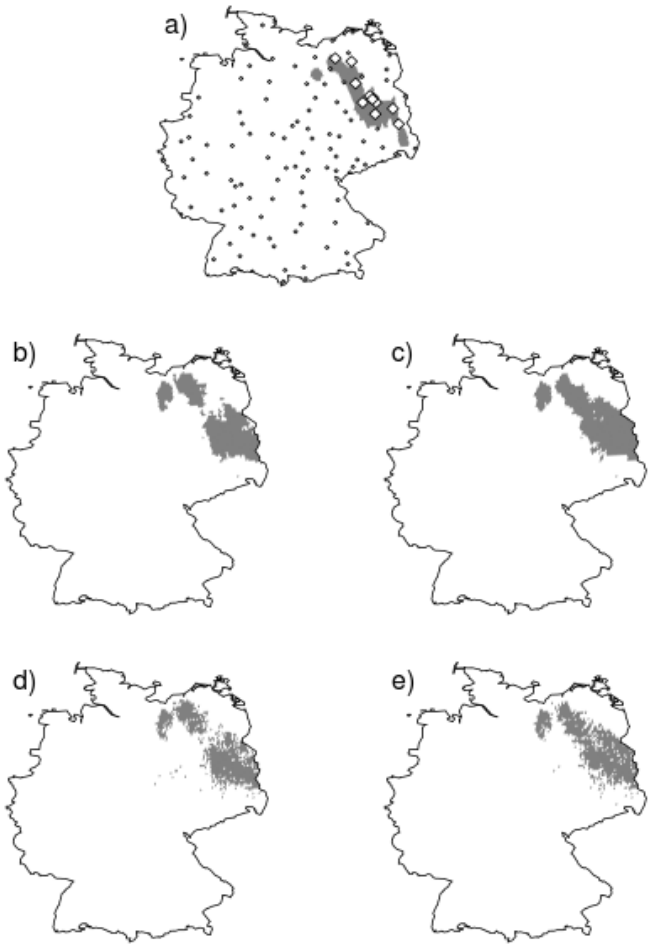

Figure 10.13.: Same as Fig. 9.5 but for occurrence/non-occurrence of precipitation $\geq$ 2.5 mm/h on July 4th, 2011 at 08 UTC.

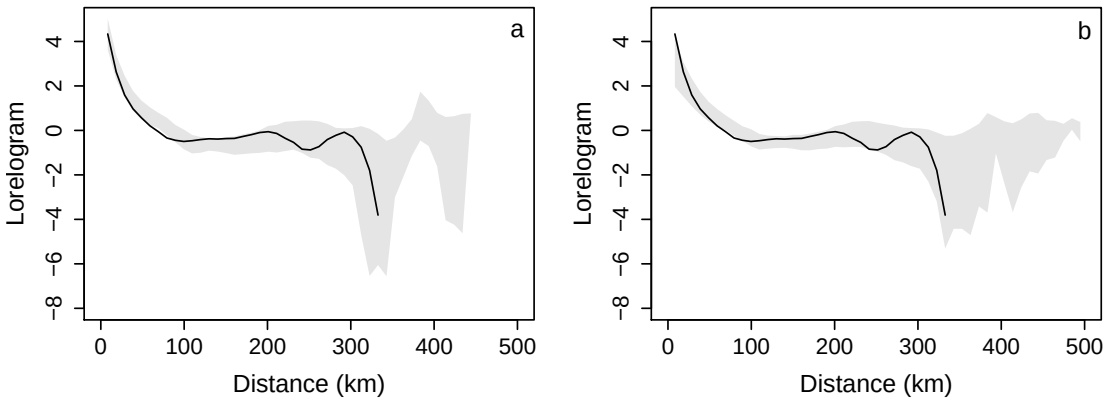

Figure 10.14.: Same as Fig. 9.6 but for precipitation $\geq$ 2.5 mm/h on July 4th, 2011 at 08 UTC.

Beside the fact that HIRAIN overestimates the rainy area, the spatial structure of precipitation $\geq$ 2.5 mm/h can be considered as one realisation of HIRAIN up to the distance of approximately 330 km (Figure 10.14). The difference between THR and BNS at the nearby distances is once more emphasized. THR has better performance than BNS in reproducing the log odds ratio of about 4 shown by radar.

The realisations of $\boldsymbol{y}^p$ for the 7.0 mm/h threshold for June 22nd, 2011 at 16 UTC largely overestimate the rainy area observed by radar (Figure 10.15). BIAS is 5.23 for THR and 5.15 for BNS (table 10.5). This indicates that the number of grid points with precipitation predicted by HIRAIN is about 5 times greater than the number of grid points where precipitation is observed. The training stations and radar did not detect precipitation $\geq$ 7.0 mm/h in the same region. The rainy area detected by radar is much smaller than the one covered by the stations. As seen for other precipitaiton events, in such cases the disagreement between the predictions and radar does not indicate poor performance of HIRAIN, but it is due to the fact that radar and stations do not agree.

The spatial dependence structure is well reproduced by both methods (Figure 10.16), except for the underestimation of the log odds ratio by BNS at nearby locations. Both methods slightly overestimate the spatial dependence at distances between 100 km and 170 km, approximately. However, this is a particularity of

Table 10.5.: Categorical statistics for precipitation $\geq$ 7.0 mm/h on June 22nd, 2011 at 16 UTC.

Statistics	THR	BNS
BIAS	5.23	5.15
POD	0.56	0.48
FAR	0.89	0.90
Log OR	2.25	1.88
ETS	0.08	0.06
HK	0.44	0.35

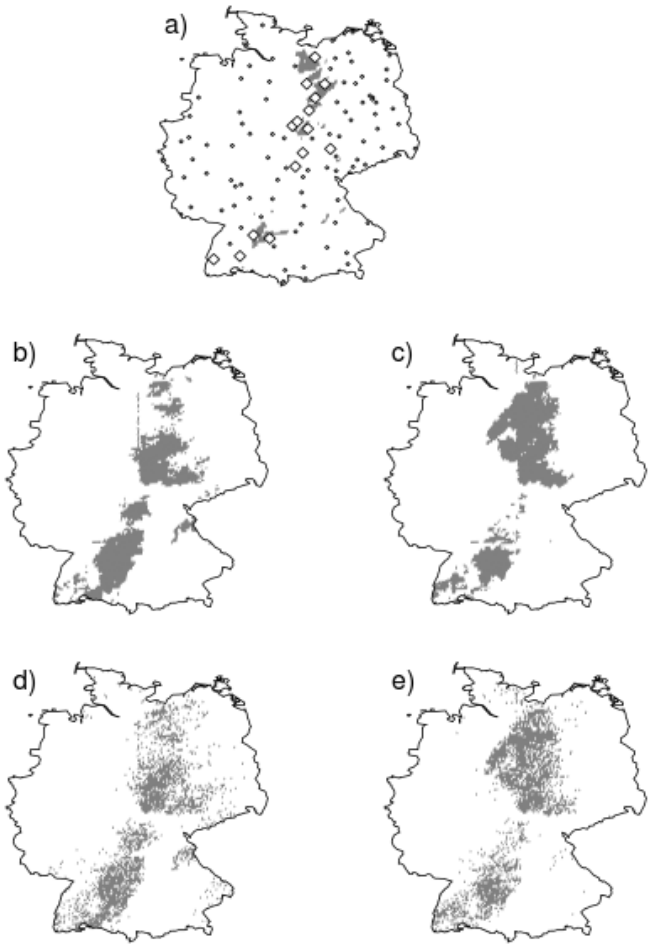

Figure 10.15.: Same as Fig. 9.5 but for occurrence/non-occurrence of precipitation $\geq$ 7.0 mm/h.

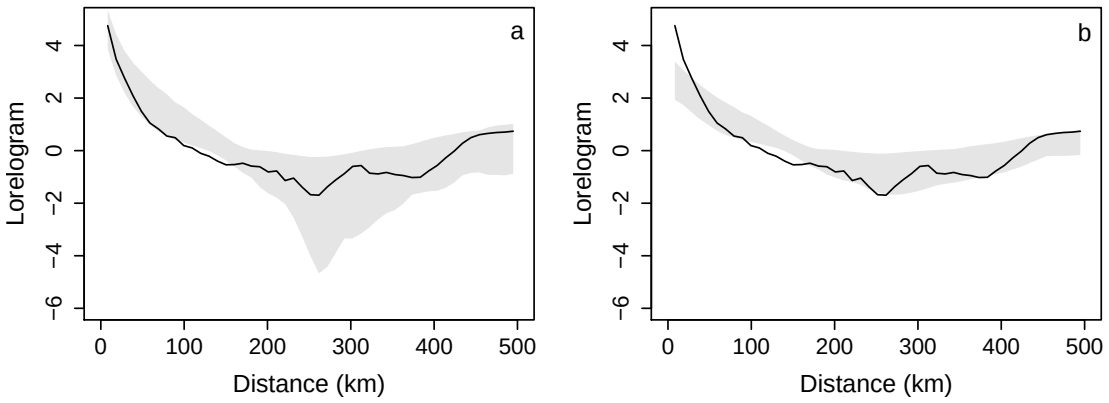

Figure 10.16.: Same as Fig. 9.6 but for precipitation $\geq$ 7.0 mm/h.

this precipitation case and it is more pronounced for BNS.

The predictions of $\boldsymbol{y}^p$ for July 4th, 2011 at 08 UTC also overestimated the area with precipitation $\geq$ 7.0 mm/h observed by radar (Figure 10.17). The categorical statistics (table 10.6) show high values of BIAS, low POD and high FAR. ETS is nearly zero and indicates that correct predictions might often happen by chance. As for the previous case, the disagreement between radar and HIRAIN is caused due to the fact that radar and the training stations do not agree. Although the predictions do not reproduce the radar field, they are in agreement with the validation stations (not shown).

The lorelograms (Figure 10.18) illustrate the features presented by the realisations. There is no spatial dependence in the structure of precipitation $\geq$ 7.0 mm/h observed by radar at distances larger than 200 km, whereas the realisations show spatial dependence up to 350 km for THR and approximately 420 km for BNS. Except for the difference between THR and BNS at nearby distances, the realisations reproduce the radar spatial structure only up to approximately 130 km in both methods.

Table 10.6.: Categorical statistics for precipitation $\geq$ 7.0 mm/h on July 4th, 2011 at 08 UTC.

Statistics	THR	BNS
BIAS	4.6	4.25
POD	0.33	0.30
FAR	0.93	9.93
Log OR	1.31	1.30
ETS	0.04	0.04
HK	0.21	0.19

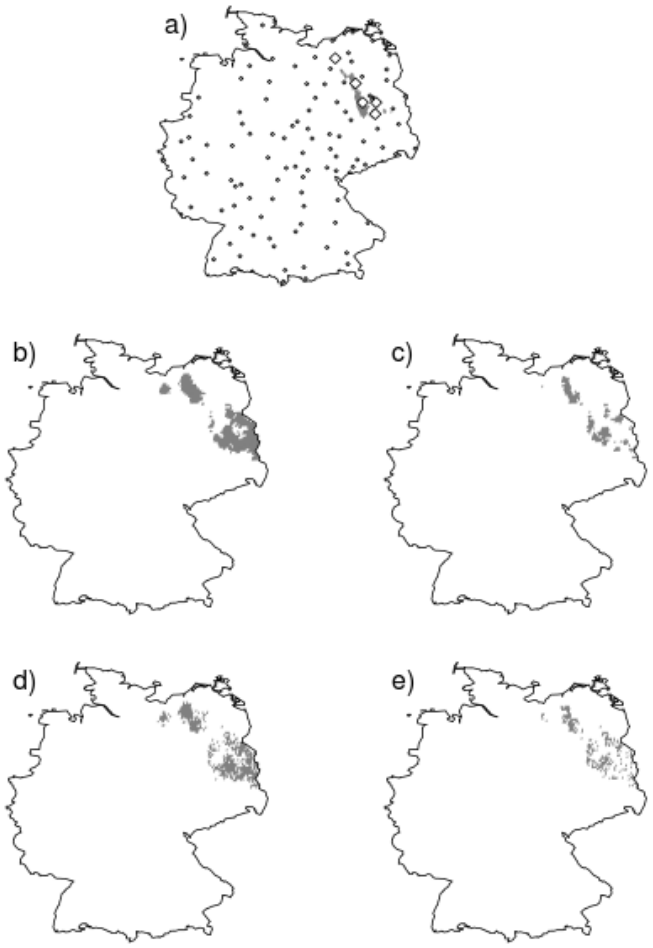

Figure 10.17.: Same as Fig. 9.5 but for occurrence/non-occurrence of precipitation $\geq$ 7.0 mm/h on July 4th, 2011 at 08 UTC.

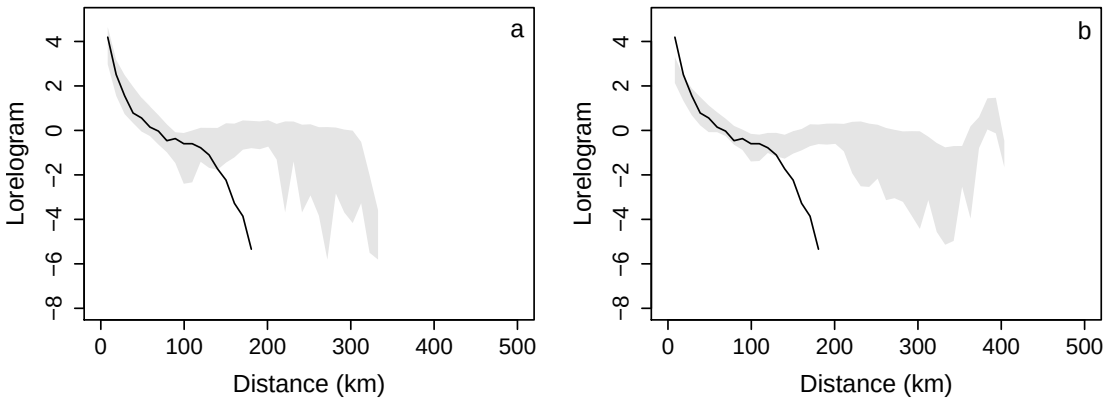

Figure 10.18.: Same as Fig. 9.6 but for precipitation $\geq$ 7.0 mm/h on July 4th, 2011 at 08 UTC.

11. Precipitation Product

The final product generated with HIRAIN in this study are hourly fields of precipitation with 4 km spatial resolution in the area of Germany from December of 2010 to November of 2011.

The precipitation product is obtained by combining the fields of $\boldsymbol{y}^p$ for the thresholds of 0.1 mm/h, 0.5 mm/h, 1.0 mm/h, 2.5 mm/h, 5.0 mm/h, 7.0 mm/h and 10.0mm/h. The field $\boldsymbol{y}^p$ of one threshold is subtracted from the other to determine areas with precipitation for different classes. The subtraction of the field of $\boldsymbol{y}^p$ for precipitation $\geq$ 0.5 mm/h from the field $\boldsymbol{y}^p$ for precipitation $\geq$ 0.1 mm/h determines the regions with precipitation amount ranging from 0.1 mm/h to 0.5 mm/h. The same logic is adopted for the other thresholds. Thus, fields of precipitation are produced in the classes between 0.1 mm/h and 0.5 mm/h, 0.5 mm/h and 1.0 mm/h, 1.0 mm/h and 2.5 m/h, 2.5 mm/h and 5.0 mm/h, 5.0 mm/h and 7.0 mm/h, 7.0 mm/h and 10 mm/h and $\geq$ 10.0 mm/h. This procedure is applied hourly to the entire period.

In order to illustrate the product obtained with HIRAIN, simulations are shown for the cases on June 22nd, 2011 at 16 UTC (Figure 11.1) and on July 04th, 2011 at 08 UTC (Figure 11.3). In both figures two realisations out of 10, obtained with THR are shown in (b) and (c) and two realisations from BNS are in (d) and (e), whereas (a) presents the corresponding radar field. Radar precipitation is ploted based on the same precipitation classes, for which the realisations are generated, i.e., between 0.1 mm/h and 0.5 mm/h, 0.5 mm/h and 1.0 mm/h, and so forth.

The evaluation is performed against radar data by the means of FSS. For that, spatial scales ranging from 0 to 10 are adopted. In the former, the simulated fields and radar are compared pointwise. From the scale of 1 to 10, an area surrounding each grid point is delimited, for which simulations and radar are compared.

Figures 11.2 and 11.4 show the FSS for the precipitation thresholds of 0.1 mm/h, 0.5 mm/h, 1.0 mm/h, 2.5 mm/h, 5.0 mm/h and 7.0 mm/h for June 22nd,

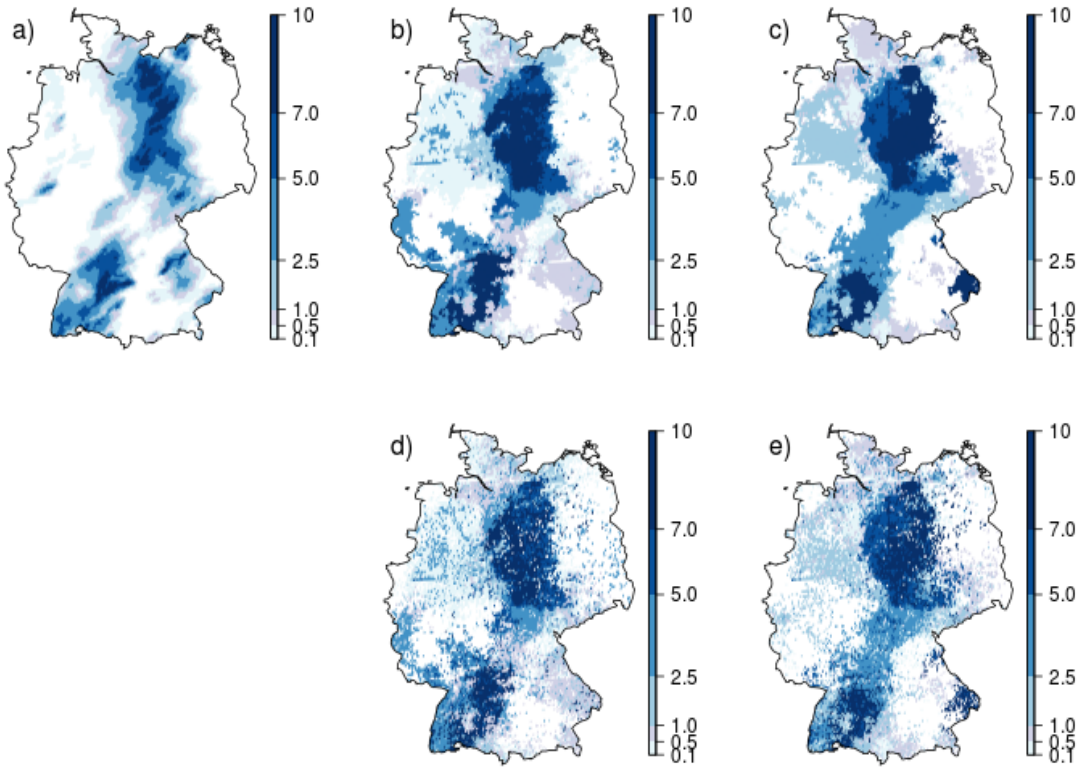

Figure 11.1.: (a) Precipitation in mm/h on June 22nd, 2011 at 16 UTC given by (a) radar, (b,c) two realisations from THR and (d,e) two realisations from BNS.

2011 at 16 UTC and for July 04th, 2011 at 08 UTC, respectively. The measure given in the x axes is the size in km^2 of the area corresponding to the scale ranging from 0 to 10.

FSS increases as the area/scale increases for both precipitation cases and for all thresholds. The lowest FSS is found when the simulations are compared with radar pointwise. The best performance of HIRAIN for all areas/scales is for occurrence/non-occurrence of precipitation ≥ 0.1 mm/h (Figures 11.2 a) and 11.4 a)). The performance diminishes as the threshold increases. The latter has already been verified in the evaluation presented in chapters 8 to 10.

Precipitation ≥ 0.1 mm/h on June 22nd, 2011 at 16 UTC (Figure 11.2 (a)) show FSS ranging between 0.77 and 0.99 as the size of the area varies from 0 to 120 km^2 for the first realisation obtained with THR. The second realisation shows

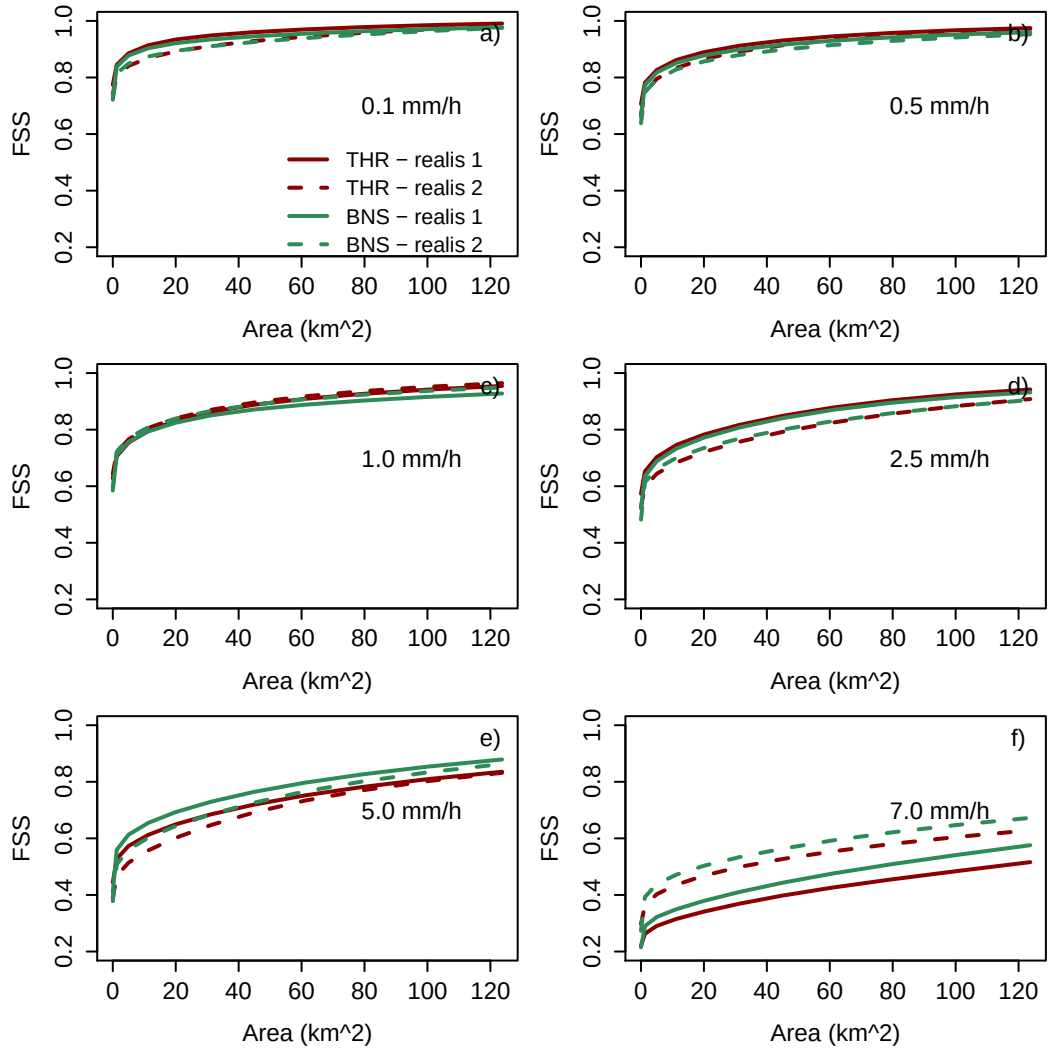

Figure 11.2.: FSS for the two realisations obtained with THR and BNS shown in Fig. 11.1 for the thresholds of (a) 0.1 mm/h (b) 0.5 mm/h (c) 1.0 mm/h, (d) 2.5 mm/h, (e) 5.0 mm/h and (f) 7.0 mm/h.

similar behaviour with FSS between 0.74 and 0.98. BNS presents FSS ranging between 0.72 and 0.97 for both realisations. Although the less homogeneous spatial structure produced by BNS, the difference in the FSS between THR and BNS is relatively small.

Occurrences of precipitation $\geq$ 0.5 mm/h and $\geq$ 1.0 mm/h do not present considerable difference in the FSS between both realisations neither for THR nor for

BNS (Figure 11.2 b),c)). Precipitation $\geq$ 0.5 mm/h shows FSS ranging from approximately 0.7 to 0.96, whereas for 1.0 mm/h the score ranges from 0.6 to 0.93.

The FSS tends to decrease for all areas/scales when occurrence/non-occurrence of precipitation $\geq$ 2.5 mm/h is considered (Figure 11.2 d)). There is a slight difference in the curves between both realisations but there are not relevant differences between THR and BNS. The first realisation shows FSS ranging from approximately 0.55 to 0.95, whereas the second shows values between 0.5 and 0.9.

There is a notable decrease in the performance of HIRAIN for the higher thresholds (5.0 and 7.0 mm/h) (Figure 11.2 e),f)). For these precipitation thresholds there is a difference in the curves between both realisations and between THR and BNS. Occurrence of precipitation $\geq$ 5.0 mm/h presents FSS ranging from 0.44 to 0.83 for the first realisation obtained with THR, whereas this decreases slightly for the second realisation with values between 0.38 to 0.83. With BNS the first realisation shows FSS ranging from 0.41 to 0.87, whereas for the second realisation the score varies from 0.38 to 0.86. Occurrence of precipitation $\geq$ 7.0 mm/h presents even lower FSS values. They range between 0.22 and 0.51 for the first realisation of THR and between 0.3 and 0.63 for the second realisation. With BNS FSS is between 0.2 to 0.57 for the first realisation and betwenn 0.27 to 0.67 for the second one.

HIRAIN is able to simulate very well the occurrence/non-occurrence of precipitation for the thresholds of 0.1 mm/h, 0.5 mm/h and 1.0 mm/h in areas $\geq$ 60 km^2, for which FSS tends to 1. For precipitation $\geq$ 2.5 mm/h, the performance tends to 100% when the simulations in areas between 80 km^2 and 100 km^2 are compared to radar. For the higher thresholds the compared areas need to be much larger than 120 km^2 for the maximum performance of HIRAIN (not shown in the figures).

The FSS computed for precipitation on July 04th, 2011 at 08 UTC (Figure 11.4) shows in general similar characteristics to those found previously, however, FSS increases faster with the increase of the area/scale for precipitation $\geq$ 5.0 mm/h and $\geq$ 7.0 mm/h than in the previous case (Figures 11.4 (e,f)).

HIRAIN has a good performance in reproducing occurrence/non-occurrence of precipitation $\geq$ 0.1 mm/h, with FSS ranging from 0.75 for a pointwise comparison to 0.99 for comparison in areas of about 120 km^2. This is shown by both reali-

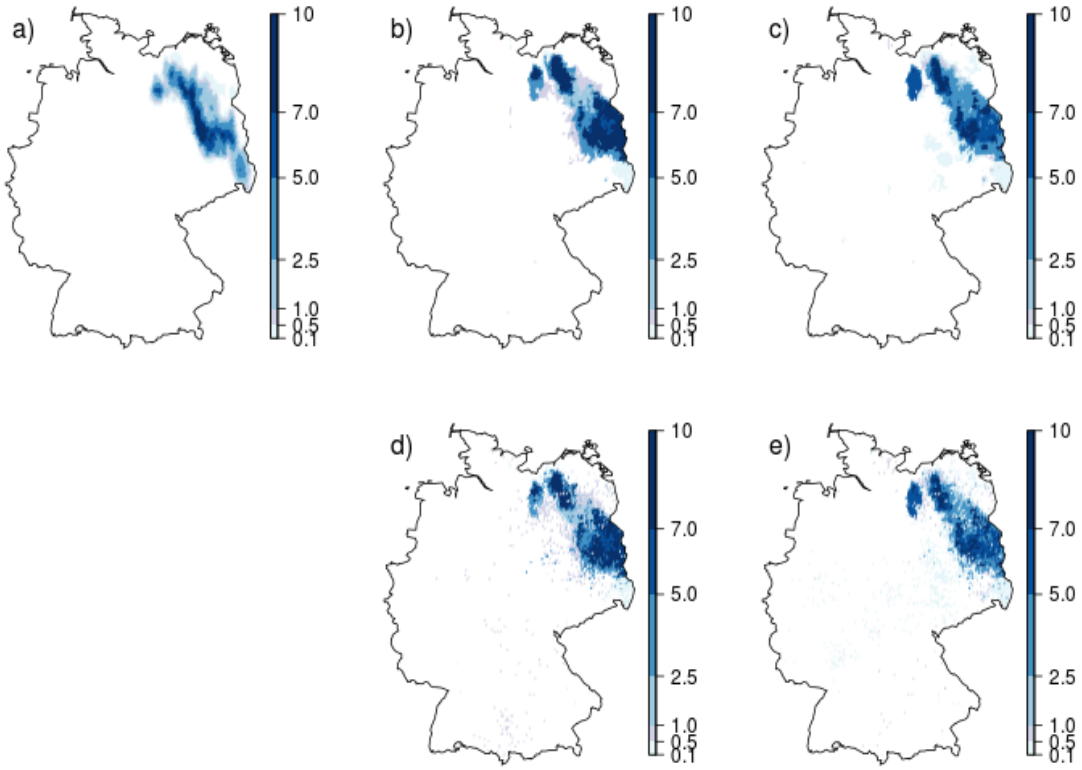

Figure 11.3.: Same as Fig. 11.1 but for rainfall event at 08 UTC on July 04th.

sations and both methods (THR and BNS).

For the thresholds of 0.5 mm/h, 1.0 mm/h and 2.5 mm/h there is a slight difference in the curves of FSS between both realisations, but not between THR and BNS. The first realisation for the 0.5 mm/h threshold shows FSS ranging from 0.7 to 0.99, while for the second realisation the scores are between 0.6 and 0.99. For precipitation $\geq$ 1.0 mm/h the FSS ranges from 0.55 to 0.99 for the first realisation and between 0.6 and 0.99 for the second. For the 2.5 mm/h threshold FSS varies from 0.52 to 0.97 and from 0.53 to 0.96 for the first and second realisation, respectively.

Differences between the performance of BNS and THR are found in the second realisation for precipitation $\geq$ 5.0 mm/h (e). FSS ranges from from 0.4 to 0.92 for THR and from 0.38 to 0.94 for BNS. The first realisation do not show differences between the curves and the skill score is between 0.3 and 0.92. For precip

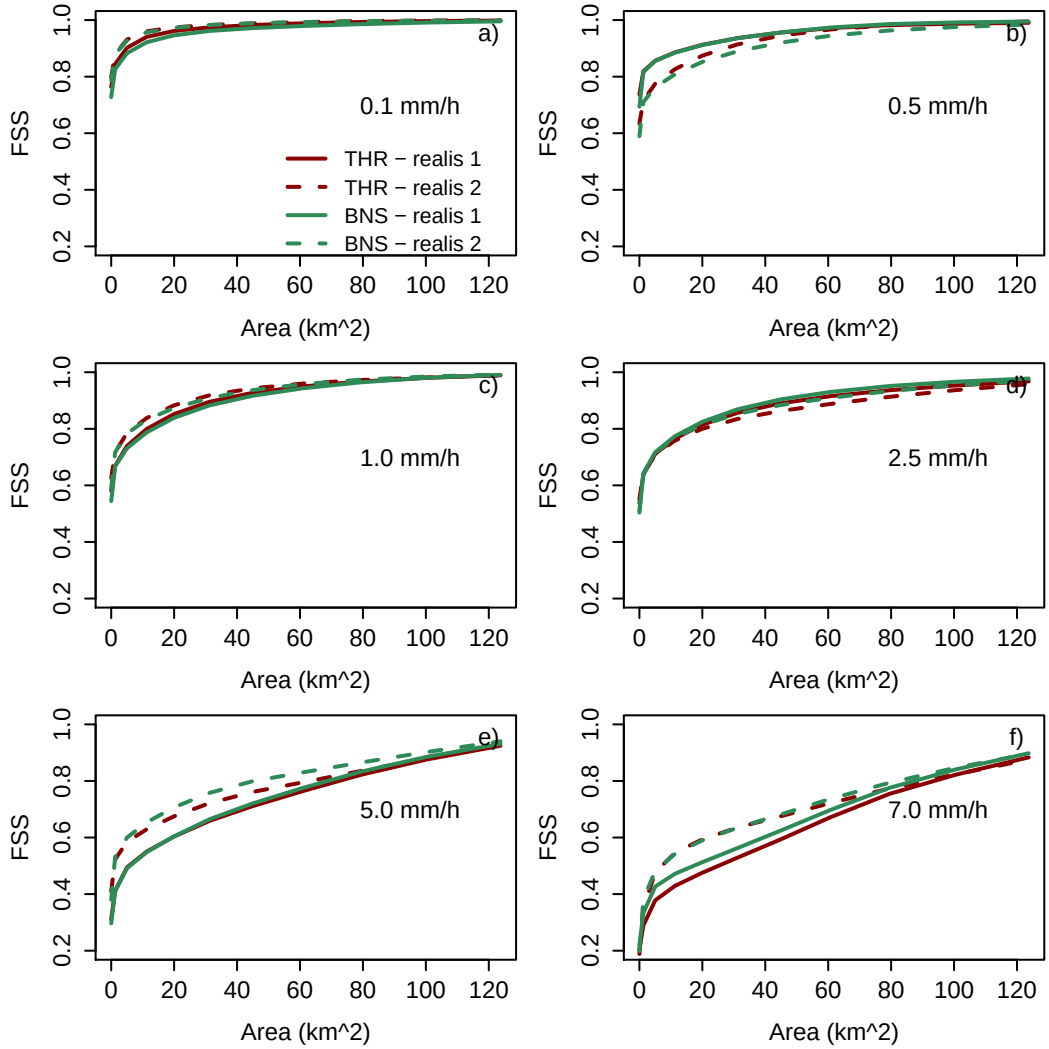

Figure 11.4.: Same as Fig. 11.2 but for the realisations in Fig. 11.3 for the rainfall event at 08 UTC on July 04th, 2011.

itation $\geq$ 7.0 mm/h, this pattern is reversed. A difference in the curves between THR and BNS is seen for the first realisation. This presents FSS ranging from 0.19 to 0.87 for THR and from 0.2 to 0.89 for BNS. The second realisation shows FSS between 0.21 and 0.89 for both methods.

12. Summary and Conclusion

Within the framework of climate research, climate retrospective analysis (reanalysis) contributes to a better understanding and monitoring of climate change. Self-consistent datasets have been produced to describe past states of the atmosphere, land surface and ocean at many research centers around the world.

The Hans Ertel Centre for Weather Research (HErZ), funded by the German Meteorological Service (DWD) set up a high resolution reanalysis based on the NWP model COSMO for Germany and Europe (Bollmeyer *et al.* 2014). The reanalysis for Germany assimilates among other variables high resolution precipitation rates through a LHN scheme. For this domain the reanalysis has been already run from 2007 to 2012 and radar precipitation rates were assimilated. For the period before 2007 radar data is not available and another set of radar-like precipitation dataset is required for the assimilation. Many precipitation products are available, however, none presented an appropriate time, spatial and temporal resolution needed for the purpose of this reanalysis. In view of that, this study proposes HIRAIN, a method to generate a more appropriate precipitation dataset for the assimilation.

Although this research is motivated by the need to produce high resolution dataset for earlier years than 2007, HIRAIN firstly generated precipitation fields for the year 2011. Due to the available radar estimates and a denser network of synoptic stations, the method could be evaluated. Once an appropriate method is available, ensembles can be generated for periods before 2007.

HIRAIN works in two steps. First, a Bayesian statistical model conditional on precipitation data from synoptic stations and satellite IRBT simulates a latent spatial Gaussian process at unobserved locations. It is assumed that this process drives the occurrence of precipitation equal or greater than a certain threshold. In a second step occurrence/non-occurrence of precipitation above the same threshold is generated given the simulated process. Many thresholds of precipitation amount are chosen, for which the occurrence/non-occurrence is simulated. A final pre-

cipitation product is obtained from the realisations of precipitation occurrence of each threshold.

The approach used in this work differs from others in many ways. Previous studies using Bayesian models have typically simulated the occurrence/non-occurrence and the amount of precipitation (Sanso, Guenni 1999; Gaudard *et al.* 1999; Berrocal *et al.* 2008). However, they have done it for lower spatial and temporal resolutions than required for this project. Here, the Bayesian statistical model is used to generate fields on a grid with 4km resolution and hourly time scale.

Predicting quantitative precipitation levels with the Bayesian model requires estimation of additional parameters. This is very difficult with the resolution used here given the limited number of available synoptic stations for longer periods back in time. Moreover, estimating additional parameters for a model to estimate precipitation for multiple years in hourly time steps would make the method overly complicated. In this scenario, the method proposed in HIRAIN is computationally tractable and more feasible to provide radar-like precipitation fields.

The advantage of a Bayesian approach is that uncertainties of the model parameters are considered for predictions, and equally probable scenarios of the spatial random process assumed to generate the observations can be simulated conditionally on the data. The data at the observation locations are honored and the spatial covariance structure of the spatial process is reproduced in each realisation. This leads to more realistic fields than those produced by interpolation methods, which tend to smooth the estimated variable. Moreover, the ability to generate ensemble of possible precipitation patterns can also be very usefull for assimilation in ensemble reanalysis.

In order to predict the occurrence/non-occurrence of precipitation exceeding a threshold, two different methodologies are applied. One approach simulates occurrence as a spatially independent Bernoulli process, which considers that the precipitation process is conditionally independent in space given the latent spatial process. This is named here Bernoulli sampling (BNS) and is consistent with the assumption that the precipitation observations are generated by a spatially independent Bernoulli process. For the neighboring distances between synoptic stations of about 5.6 km to 13.35 km, this is justified. However, because the predictions of occurrence/non-occurrence are carried out on a 4km grid, the dependence between precipitation at neighboring locations may not be ignored anymore.

A more appropriate approach to generate these fields should account for spatial dependencies in the precipitation process. That is, after accounting for the spatial dependence in the probability of precipitation (captured in the structure of S), there is likely a 'residual' spatial dependence in the occurrence/non-occurrence of precipitation above a threshold. However, the spatial dependence structure of the responses is unknown and would have to be inferred from the observations. Given the synoptic stations design, it is challenging to infer a reliable dependence structure. Moreover, proper treatment of this residual dependence in the observations would require estimation of further parameters with the BSGLM. Hence, additional MCMC simulations would be required, making the estimation procedure more time consuming. Including this dependence would greatly complicate inference and prohibit the use of 'geoRglm'.

For this reason, in a second approach occurrence/non-occurrence is generated by thresholding the realisations of the spatial Gaussian process. This is called the thresholding method (THR). All regions above an a priori defined threshold are considered to be regions with precipitation, whereas no precipitation is assigned otherwise. In this approach, it is implicitly assumed that the precipitation process is fully determined by the latent Gaussian process. This is inconsistent with the assumption in the BSGLM, however, the spatial dependence that is considered in the process might nevertheless provide more realistic fields.

The performance of HIRAIN is evaluated for each threshold of precipitation amount separatelly. For the lowest threshold that comprehends all precipitation intensities, namely precipitation $\geq$ 0.1 mm/h, the simulations showed to be sensitive to the spatial distribution of the observations, the number of observations with precipitation and the degree of relation between ground observations and satellite information.

Through the evaluation of the simulations against a second dataset of synoptic stations, high performance is achieved when larger numbers of stations with precipitation are present, more than 25 stations with precipitation are wide distributed and well distributed over the country or less than 25 rainy stations, however, concentrated over one region were considered and when there are high negative correlation coefficient between precipitation and IRBT. Any change in these configurations decreases the model's performance.

The performance is reduced when a small number of stations with precipitation

is wide distributed over the country. When very few observations present large distances among them, it is very difficult for the model to infer a reliable scale parameter. Beside the fact that the predictions are not reliable themselves, the time required for the MCMC simulations to converge increases considerably. The performance is the lowest, when this condition meets weak relationship between precipitation and IRBT.

Every month in the period from December 2010 to November 2011 presents cases with good and bad performance. However, in general the simulations agree better with the validation stations during winter than the other seasons, especially summer. This difference indicates that the favorable conditions for high performance of HIRAIN are present more often during winter than during summer.

With the objective to simulate radar-like datasets, the ability of HIRAIN to capture the spatial dependence structure detected by radar is investigated. To this end, the predicted occurrence/non-occurrence of precipitation exceeding a threshold is validated against radar data. Categorical statistics and the lorelogram show that THR agree better with radar than BNS. Through the lorelograms there is a clear difference between the log odds ratio of THR and BNS at neighboring locations. THR estimates very well the log odds ratio of radar, whereas BNS underestimates it at those locations. This difference is also seen in the fields of occurrence/non-occurrence. The spatial structure produced by BNS is more punctual than that produced by THR. The latter shows more homogeneous and realistic fields.

For the higher thresholds of precipitation amount, it becomes more difficult to simulate the fields. The reduced number of stations with precipitation above a selected threshold and the unfavorable spatial distribution play the most important role for these simulations. The first threshold for which this difficulty appears is precipitation $\geq$ 2.5mm/h. The estimation of the scale parameter is more difficult and therefore it takes also longer to tune the model. This increases the time the Bayesian model needs to run. Similar degree of difficulty is found for precipitation amounts between 2.5 mm/h and 7.0 mm/h. For amounts larger than 7.0 mm/h this difficulty increases even more, when too few stations detect precipitation and the relation between precipitation and IRBT is very weak. However, since larger precipitation amounts are also usually associated with convective clouds, there is a stronger relationship between station information and IRBT, which tends to

improve the performance of HIRAIN in this case.

The final precipitation product produced in this study is a 1-year dataset for Germany with hourly and 4 km resolutions. This product is obtained from the occur rence/non-occurrence fields of the individual thresholds. In comparison with radar the ability of HIRAIN to predict precipitation is the best for lower than for higher amounts. The performance also increases when an area surrounding the grid point under consideration is compared to the same area in the radar field. The larger the area, the better the performance.

HIRAIN can be applied to any geographical region and can be extended to further thresholds of precipitation amounts and finer resolutions. The resolution adopted in this study most appropriately matched the feasibility of the method with the purposes of the reanalysis. The final product presents better resolutions than the precipitation datasets available by the time the Germany reanalysis project started. The method is evaluated and available to produce precipitation datasets also for other time periods.

It is proposed as future work to evaluate the reanalysis for Germany when these generated fields are assimilated in the COSMO model. Beside that, it is also suggested further implementations in HIRAIN to improve the prediction's performance. The inclusion of other covariates besides IRBT might improve predictions, especially for precipitation cases as seen in category C, where there is not enough precipitation information and thus, it is challenging for the BSGLM to estimate the scale parameter ϕ.

The running time of HIRAIN should be optimized through parallelisation. This will allow for further implementations in the method as the use of more precipitation thresholds. Hence, high resolution precipitation fields can be generated for finer classes of precipitation amount.

The realisations obtained with HIRAIN provide also valuable information of precipitation uncertainties that plays an important role in ensemble reanalysis (Bach *et al.* 2016). In this scenario, it is also suggested as future work to increase the number of realisations obtained with HIRAIN, once HIRAIN has been parallelized. The larger the number of realisations, the more accurate is the information about the precipitation uncertainties. This allows for estimating the sensitivity of the reanalysis to the uncertainties in the observations and might contribute for better performance of ensemble reanalyses.

Appendix

Table of appendices

A. Prior probability distribution

As seen in chapter 4, a Bayesian hierarchical model is given by a data model, a process model and a parameter model. In this chapter the focus will be only on the parameter model. This is given through a probability distribution containing prior information about the parameters before the observations are collected. It is described here which prior information is provided and which considerations are assumed in the BSGLM.

The parameters that require a prior probability distribution in the model are $\boldsymbol{\beta}$, σ^2 and ϕ. The former is a vector that contains the regression coefficients β_0 and β_1 that provide the relationship between the covariate and the occurrence/non-occurrence of precipitation. The parameter σ^2 is the variance of the spatial Gaussian process and ϕ is the scale that indicates the distance up to which occurrences/-non-occurrences are correlated in space.

Usually non-informative priors are used when there is no previous information about the parametes. In such cases, all values for a given range have the same probability to occur and the model obtains most of the information from the observed data. When some previous information is known a most appropriate prior can be provided to the model.

In this study it is assumed that the regression coefficients in $\boldsymbol{\beta}$ are uniformly distributed over $(-\infty, +\infty)$, whereas informative priors are used for the parameters σ^2 and ϕ. Given the fact that the number of observations is reduced, it is challenging to estimate the parameters only from the data. Providing additional information through the priors facilitates, therefore, the estimation.

Prior information of σ^2 and ϕ is obtained from the climatology, which is based on hourly precipitation data from the validation stations in the period from 2004 to 2010 and from 2012 to 2014. Since the period of study is from December of 2010 to November of 2011, data in the period from January to November of 2011 as well as from December of 2010 is not used.

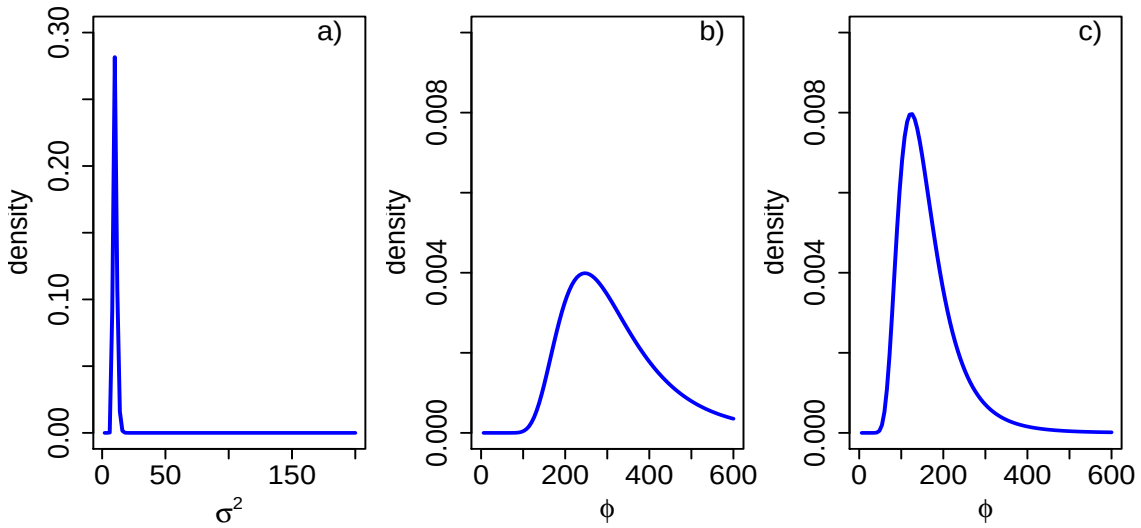

Figure A.1.: (a) Inverse scaled χ^2 distribution with df=100 and scale=10. (b) Inverse gamma distribution with shape=7.1 and scale=2000. (c) Inverse gamma distribution with shape=7.1 and scale=1000. The parameter ϕ is given in km.

Precipitation is converted to binary information. An one is assigned to the stations and to the hours with precipitation ≥ 0.1 mm/h and zero elsewhere. Lorelograms are computed for every hour with valid data. It is assumed that the spatial dependence structure of occurrence/non-occurrence of precipitaiton is described by an exponential covariance function. Thus, the function $\sigma_*^2 exp(-h/\phi_*)$ is fitted to the lorelograms. The h is the distance between two stations in space and the parameters σ_*^2 and ϕ_* are estimates of σ^2 and ϕ. Histograms of these estimates are generated for each month, separately. An scaled inverse χ^2 and an inverse gamma distribution is fitted to the histograms of σ_*^2 and ϕ_*, respectively.

The fitted inverse gamma distributions provided very high probabilities of occurrence for small values of ϕ ($\phi \leq 100$ km). When these distributions are used as prior distributions in the BSGLM, the distance up to which the outcomes are correlated in space is very small. This generates fields of probability of precipitation with a more punctual structure than that of the precipitation fields. The BSGLM presented poor performance even for cases with precipitation covering large spatial areas. Therefore, in order to predict more realistic probability fields, the inverse gamma distribution for every month is centered at larger values and this is thus used as prior for ϕ.

Due to the similarities of the probability distributions among different months, an inverse gamma with shape of 7.1 and scale of 2000 is adopted for the period from May to September of 2011 (M-S), and for the months from October to November of 2011, January to April of 2011 and December of 2010 (O-A) this distribution has the same shape, however with scale of 1000 (Figure A.1 (b) and (c)). During M-S the curve is centered at about 250 km and distances below 100 km present probability about zero. For the period O-S the gamma distribution is centered at about 130 km and distances lower than 50 km present probability about zero.

The same logic is adopted to estimate a probability distribution to be used as prior for σ^2. Given the fact that, the number of training stations used in the study is reduced, that $\boldsymbol{\beta}$ uses a non-informative prior and ϕ uses wide dispersion distributions, the prior of σ^2 needed to have narrower dispersion than the formers to enable the MCMC simulations to converge. The chains would not converge for many precipitation cases otherwise. Based on that, the scaled inverse χ^2 distribution is chosen to be very narrow and centered at the value for which the histograms showed the highest frequency of occurrence. Since there is not much difference among the curves for different months, the scaled inverse χ^2 distribution with 100 degrees of freedom and scale of 10 is adopted as prior from December of 2010 to November of 2011 (Figure A.1 a)).

The prior distributions adopted to simulate probability of precipitation ≥ 0.1 mm/h are also used for the higher thresholds of precipitation amount. Since the number of stations with precipitation occurrence decreases as the threshold increases, it becomes harder to obtain estimates ϕ_* and σ_*^2 from the climatology.

B. MCMC convergence diagnostics

As already described in chapter 5, MCMC simulations are performed with the BS-GLM to approximate the model's parameter ϕ and the spatial Gaussian process $\boldsymbol{s}^o$ at the observation locations $\boldsymbol{r}^o$. In order to assure that the samples of ϕ and $\boldsymbol{s}^o$ are drawn from a stationary distribution, convergence diagnostics are required. To illustrate the diagnostics used in this study, an example is shown for the simulations of probability of precipitation ≥ 0.1 mm/h on June 22, 2011 at 22 UTC. For simplicity the results are presented for 8 of 121 observation locations.

The tests are applied to the chains of ϕ and $\boldsymbol{s}^o$ obtained from 15.000 iterations with no burn.in. First, the trace and autocorrelation plots are visually investigated (Figures B.1 and B.2). The trace plots indicate good mixing and convergence seems to be reached in the first few iterations. The autocorrelation shows how nearby samples in the chain are correlated and it is useful for defining an appropriate value for thin in the model. The autocorrelation graphs show that 100 is a good value for thin, since after about 100 iterations the autocorrelation decreases and tends to zero.

Beside the visual investigation, the Heidelberg and Welch's diagnostics as well as the Geweke's diagnostic are applied. The Heidelberg and Welch's diagnostics are separated in two tables: table B.1 shows the stationarity test and table B.2 presents the halfwidth test.

In the stationarity test the samples of ϕ and $\boldsymbol{s}^o$ passed the test, except for one location (not shown). The number of iterations to dismiss and the p-value from the Cramer-von-Mises statistic test is shown in table B.1. The test is performed at the significance level of 95%. The rejected case presented p-value equal 0.0173 (not shown). The start iterations shown in table B.1 suggest the most appropriate value for burn.in.

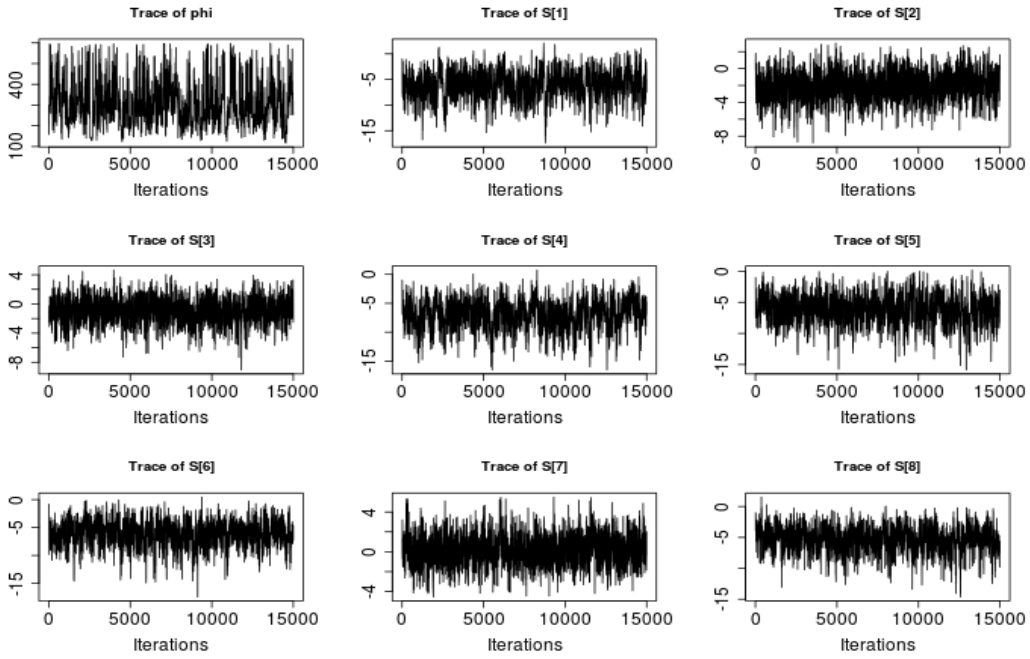

Figure B.1.: Trace plots of ϕ and the Gaussian process at 8 observation locations simulated for precipitation ≥ 0.1 mm/h on June 22, 2011 at 22 UTC.

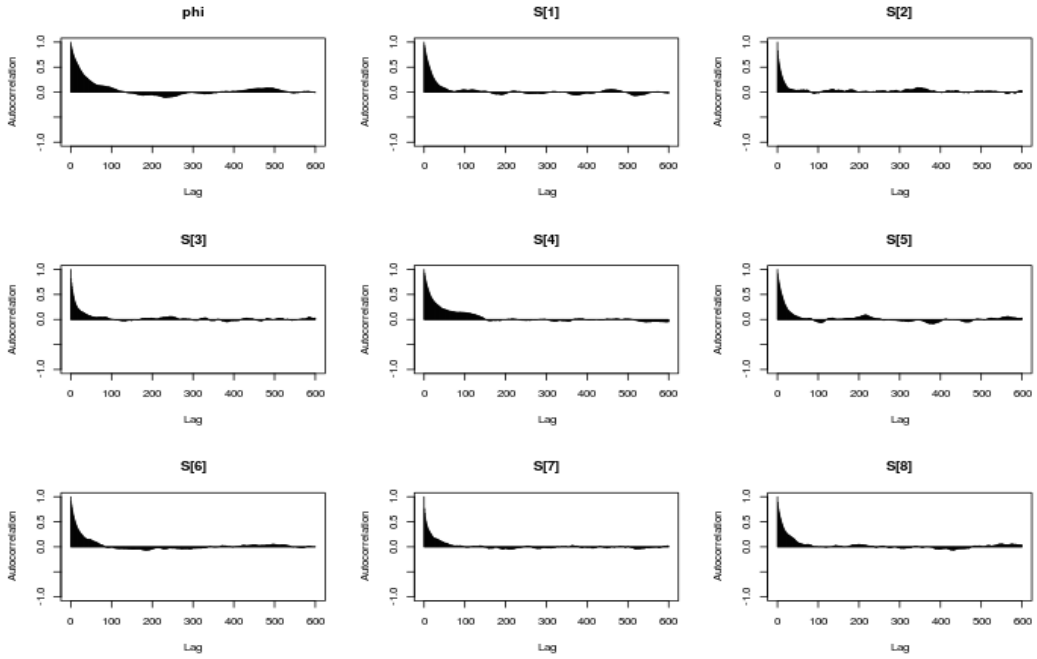

Figure B.2.: Same as Fig. B.1 but for autocorrelation of the samples.

In the second part, the halfwidth test verifies at the significance level of 95% whether the chain size that passes the stationary test is long enough to estimate the mean value. The target value to which the ratio of both means is compared amounts to 0.1. Table B.2 shows the mean value and the halfwidth. From the 113 locations not shown in the tables, 104 passed this test.

The z-scores and p-values computed for the Geweke's convergence diagnostics are shown in Table B.3. The test is performed at the significance level of 95%. From the 113 observation locations not shown in the table, 95 passed the test.

Given the fact that, it is chalenging to estimate the model's parameters from the reduced number of training stations used in this study, the following approach is adopted to determine convergence.

If the chain of ϕ and the chains of $\boldsymbol{s}^o$ for 90% of the observation locations passed the tests, those chains that did not passed them, are again visually investigated. If the trace plots at those locations show visually to be approximately stationary, the samples are stored for all locations. Otherwise the BSGLM needs to be tuned again.

The diagnostics for the other thresholds of precipitation amount present very similar results (not shown). However, it is important to highlight that the higher the threshold, the more difficult it is to tune the model.

It was determined that the MCMC simulations should perform 15.000 iterations, with a burn.in of 5.000 and store only every 100th sample. These numbers are obtained from the convergence diagnostics applied to every threshold of precipitation amount used in this study.

Table B.1.: Heidelberg and Welch's Stationarity Test. Precipitation $\geq$ 0.1 mm/h

	stationarity test	start iteration	p-value
phi	passed	1	0.3694
S[1]	passed	1	0.3252
S[2]	passed	4501	0.1506
S[3]	passed	1	0.4134
S[4]	passed	1	0.4704
S[5]	passed	1	0.1527
S[6]	passed	1	0.2606
S[7]	passed	1	0.1072
S[8]	passed	1	0.0602

Table B.2.: Heidelberg and Welch's Halfwidth Test. Precipitation $\geq$ 0.1 mm/h

	halfwidth test	mean	halfwidth
phi	passed	303.4311	12.672
S[1]	passed	-6.1118	0.256
S[2]	passed	-1.9184	0.126
S[3]	failed	-0.9651	0.111
S[4]	passed	-6.9516	0.258
S[5]	passed	-6.1610	0.196
S[6]	passed	-6.1642	0.198
S[7]	failed	0.2300	0.106
S[8]	passed	-5.3149	0.170

Table B.3.: Geweke's Test. Precipitation $\geq$ 0.1 mm/h

	convergence test	z-score	p-value
phi	passed	1.46097	0.072
S[1]	passed	-1.50007	0.066
S[2]	failed	-3.06146	0.001
S[3]	passed	-0.36144	0.358
S[4]	passed	-0.41824	0.337
S[5]	failed	1.66405	0.048
S[6]	passed	0.48878	0.312
S[7]	passed	-0.02855	0.488
S[8]	failed	2.51832	0.005

C. Precipitation cases

The precipitation cases selected for the categories A, B and C described in chapter 8 are presented in the following tables.

Table C.1.: Precipitation cases in the categories A, B and C.

Category A	Category B	Category C
2010/12/01 18UTC	2010/12/02 19UTC	2010/12/04 21UTC
2010/12/05 16UTC	2010/12/06 24UTC	2010/12/09 21UTC
2010/12/07 19UTC	2010/12/07 18UTC	2010/12/13 23UTC
2010/12/10 21UTC	2010/12/14 13UTC	2010/12/19 09UTC
2011/01/06 08UTC	2011/01/08 02UTC	2011/01/08 09UTC
2011/01/06 16UTC	2011/01/08 23UTC	2011/01/11 16UTC
2011/01/13 09UTC	2011/01/15 24UTC	2011/01/18 15UTC
2011/01/25 14UTC	2011/01/18 06UTC	2011/01/19 20UTC
2011/02/04 08UTC	2011/02/01 23UTC	2011/02/03 05UTC
2011/02/11 05UTC	2011/02/04 05UTC	2011/02/12 24UTC
2011/02/13 05UTC	2011/02/05 09UTC	2011/02/13 10UTC
2011/02/27 01UTC	2011/02/15 05UTC	2011/02/15 09UTC
2011/03/14 05UTC	2011/03/10 02UTC	2011/03/09 20UTC
2011/03/16 20UTC	2011/03/10 23UTC	2011/03/09 23UTC
2011/03/17 12UTC	2011/03/13 21UTC	2011/03/13 22UTC
2011/03/17 22UTC	2011/03/16 11UTC	2011/03/16 22UTC
2011/04/03 20UTC	2011/04/03 08UTC	2011/04/04 13UTC
2011/04/04 01UTC	2011/04/05 24UTC	2011/04/27 06UTC
2011/04/04 06UTC	2011/04/14 01UTC	2011/04/27 18UTC
2011/04/12 11UTC	2011/04/26 07UTC	2011/04/29 17UTC
2011/05/14 19UTC	2011/05/12 11UTC	2011/05/12 14UTC
2011/05/15 13UTC	2011/05/19 24UTC	2011/05/15 17UTC
2011/05/22 16UTC	2011/05/21 15UTC	2011/05/22 13UTC
2011/05/31 21UTC	2011/05/28 22UTC	2011/05/31 11UTC

Table C.2.: Cont. table C.1

Category A	Category B	Category C
2011/06/06 16UTC	2011/06/05 02UTC	2011/06/05 16UTC
2011/06/08 02UTC	2011/06/06 23UTC	2011/06/06 18UTC
2011/06/16 18UTC	2011/06/22 12UTC	2011/06/18 10UTC
2011/06/22 16UTC	2011/06/30 13UTC	2011/06/29 18UTC
2011/07/07 20UTC	2011/07/04 08UTC	2011/07/09 19UTC
2011/07/10 20UTC	2011/07/09 23UTC	2011/07/10 18UTC
2011/07/13 17UTC	2011/07/10 16UTC	2011/07/18 12UTC
2011/07/13 22UTC	2011/07/27 20UTC	2011/07/25 01UTC
2011/08/06 20UTC	2011/08/11 03UTC	2011/08/03 18UTC
2011/08/08 12UTC	2011/08/15 03UTC	2011/08/05 15UTC
2011/08/14 16UTC	2011/08/18 21UTC	2011/08/08 18UTC
2011/08/24 20UTC	2011/08/21 13UTC	2011/08/24 18UTC
2011/09/04 20UTC	2011/09/03 24UTC	2011/09/04 16UTC
2011/09/08 07UTC	2011/09/11 15UTC	2011/09/07 17UTC
2011/09/11 18UTC	2011/09/19 09UTC	2011/09/11 24UTC
2011/09/18 09UTC	2011/09/27 09UTC	2011/09/17 06UTC
2011/10/06 20UTC	2011/10/12 24UTC	2011/10/06 08UTC
2011/10/07 13UTC	2011/10/18 12UTC	2011/10/07 09UTC
2011/10/08 15UTC	2011/10/19 09UTC	2011/10/19 15UTC
2011/10/09 22UTC	2011/10/26 16UTC	2011/10/26 05UTC
2011/11/26 01UTC	2011/11/26 24UTC	2011/11/25 17UTC
2011/11/27 15UTC	2011/11/27 14UTC	2011/11/25 22UTC
-	2011/11/29 24UTC	2011/11/26 02UTC
-	2011/11/30 06UTC	2011/11/30 03UTC

D. Significance Test

A two sided significance t-test is applied to the correlation coefficient between precipitation amount at the training synoptic stations and IRBT.

The following tables show the summary of the tests for the precipitation cases in chapters 8 and 10.

Table D.1.: Category A. Precipitation $\geq$ 0.1 mm/h on May 15th, 2011 at 13 UTC.

correlation coefficient	-0.159
t	-1.757
df	118
p-value	0.081
95% confidence interval	-0.329 0.020

Table D.2.: Category A. Precipitation $\geq$ 0.1 mm/h on June 22nd, 2011 at 16 UTC.

correlation coefficient	-0.303
t	-3.440
df	117
p-value	0.0008
95% confidence interval	-0.458 -0.130

Table D.3.: Category B. Precipitation $\geq$ 0.1 mm/h on December 14th, 2010 at 13 UTC.

correlation coefficient	-0.063
t	-0.688
df	119
p-value	0.492
95% confidence interval	-0.239 0.117

Table D.4.: Category B. Precipitation $\geq$ 0.1 mm/h on July 04th, 2011 at 08 UTC.

correlation coefficient	-0.673
t	-9.889
df	116
p-value	2.2e-16
95% confidence interval	-0.764 -0.564

Table D.5.: Category C. Precipitation $\geq$ 0.1 mm/h on June 05th, 2011 at 16 UTC.

correlation coefficient	-0.244
t	-2.712
df	116
p-value	0.007
95% confidence interval	-0.407 -0.066

Table D.6.: Category C. Precipitation $\geq$ 0.1 mm/h on October 19th, 2011 at 15 UTC.

correlation coefficient	-0.143
t	-1.549
df	115
p-value	0.124
95% confidence interval	-0.316 0.39

Table D.7.: Category A. Precipitation $\geq$ 2.5 mm/h on June 22nd, 2011 at 16 UTC.

correlation coefficient	-0.293
t	-3.321
df	117
p-value	0.001
95% confidence interval	-0.449 -0.119

Table D.8.: Category B. Precipitation $\geq$ 2.5 mm/h on July 04th, 2011 at 08 UTC.

correlation coefficient	-0.652
t	-9.257
df	116
p-value	1.3e-15
95% confidence interval	-0.744 -0.534

Table D.9.: Category A. Precipitation $\geq$ 7.0 mm/h on June 22nd, 2011 at 16 UTC.

correlation coefficient	-0.277
t	-3.123
df	117
p-value	0.002
95% confidence interval	-0.436 -0.102

Table D.10.: Category B. Precipitation $\geq$ 7.0 mm/h on July 04th, 2011 at 08 UTC.

correlation coefficient	-0.493
t	-6.102
df	116
p-value	1.4e-08
95% confidence interval	-0.618 -0.343

E. Spatial dependence structure

E.1. Further precipitation cases

Realisations of occurrence/non-occurrence of precipitation ≥ 0.1 mm/h obtained with THR and BNS are presented here for the precipitation cases on May 15th, 2011 at 13 UTC (category A), on December 14th, 2010 at 13 UTC (category B) and on June 05th, 2011 at 16 UTC (category C).

For the precipitation event in May radar and the training stations are not in good agreement (Figure E.1 a)). Radar estimated precipitation in a larger area than the stations. In addition to that, there is no significant correlation between precipitation and IRBT, as shown in chapter 8.

The realisations of HIRAIN underestimated the occurrence of precipitaiton ≥ 0.1 mm/h in comparison to radar (BIAS < 1) (table E.1). POD is 0.41 for THR and 0.36 for BNS on the costs of the FAR values of 0.44 and 0.48, respectively. The odds of making a hit is greater than the odds of making a false alarm in both methods. The correct predictions accounting for hits to happen by chance produce an ETS of 0.10 and 0.07 for THR and BNS, respectively. The ability to separate between occurrence and non-occurrence of precipitation is indicated by the HK of 0.18 (THR) and 0.13(BNS). The realisations shown in Figure E.1 reassure the results presented by the categorical statistics. Both methods presented poor performance in comparison to radar.

Both THR and BNS are able to reproduce the spatial dependence structure of precipitation ≥ 0.1 mm/h for larger distances than 100 km (Figure E.2). THR overestimated the spatial dependence from neighboring locations to the distance up to 100 km, whereas BNS underestimated it at neighboring locations and overestimated it from distances between 40 km and 90 km, approximately.

Table E.1.: Categorical statistics for the precipitation case on May 15th, 2011 at 13 UTC.

Statistics	THR	BNS
BIAS	0.74	0.71
POD	0.41	0.36
FAR	0.44	0.48
Log OR	0.85	0.61
ETS	0.10	0.07
HK	0.18	0.13

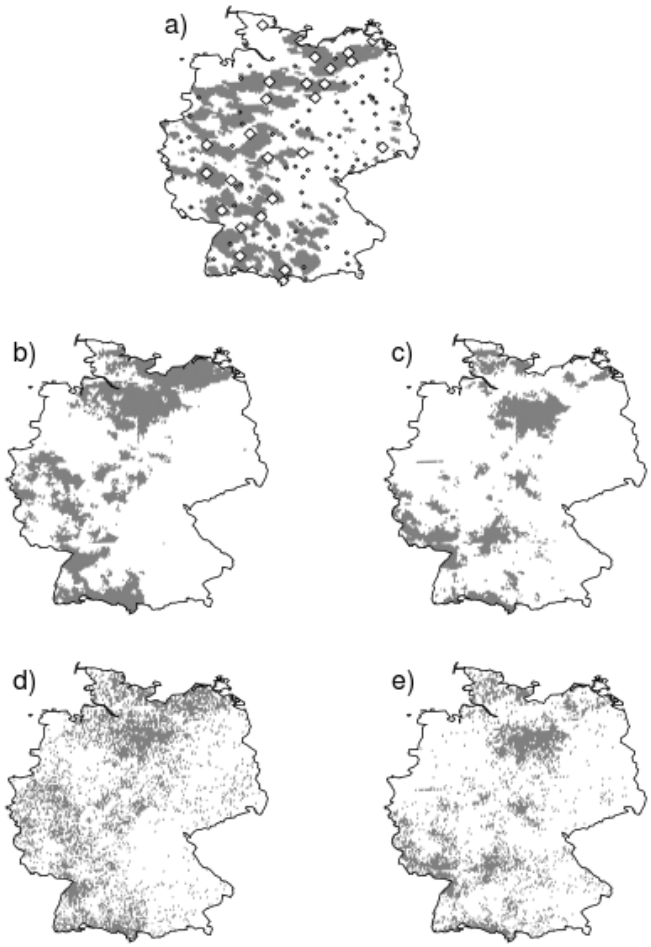

Figure E.1.: Same as Fig. 9.5 but for the precipitation case on May 15th, 2011 at 13 UTC.

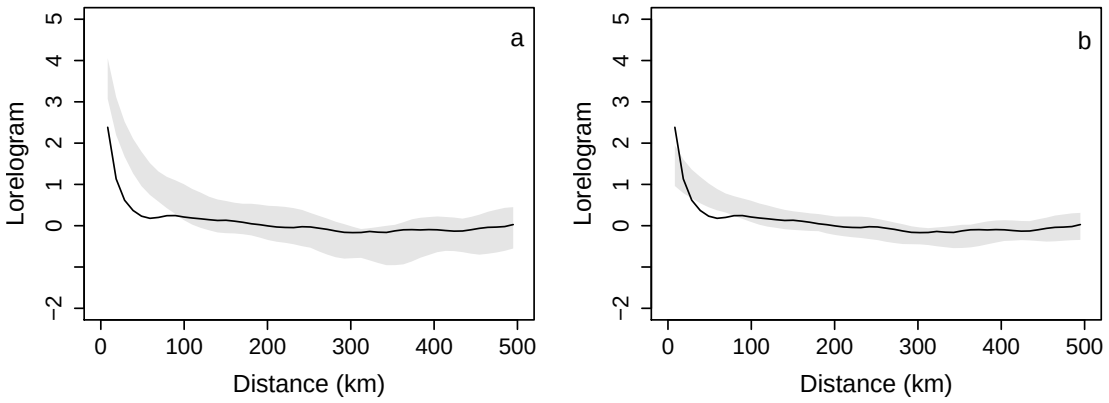

Figure E.2.: Same as Fig 9.6 but for the precipitation case on May 15th, 2011 at 13 UTC.

For the precipitation occurrence on December 14th, 2010 at 13 UTC HIRAIN shows good performance for THR and BNS (Figure E.3), although the rainy region is a bit overestimated (BIAS > 1, table E.2). HIRAIN detected 58% (THR) and 49% (BNS) of the precipitation ≥ 0.1 mm/h observed by radar, however, on the costs of the high FAR values of 63% and 65%, respectively. This result is confirmed by the visual verification when comparing the realisations to radar in Figure E.3. HIRAIN simulates a larger rainy area in the central-eastern part of Germany and predicts precipitaiton in the northeast region of the country, where radar do not observe precipitation.

There is a skill in HIRAIN in terms of the correct predictions accounting for hits to happen by chance and regarding the ability to separate between occurrence and not-occurrence of precipitation. This is shown by the values of ETS of about 0.15 (THR) and 0.12 (BNS) and of HK of 0.31 (THR) and 0.24 (BNS). The fields of $\boldsymbol{y}^p$ generated with THR are more realistic than those obtained from BNS (Figure E.3). This is reinforced by the lorelograms (Figure E.4). The spatial dependence structure is well reproduced with THR, except for distances between approximately 330 km and 380 km, which is a particularity of this precipitation case. BNS, however, underestimates the log odds ratio at neighboring locations, which is a deficiency of this method, and overestimates the spatial dependence at distances larger than 300 km.

Table E.2.: Categorical statistics for the precipitation case on December 14th, 2010 at 13 UTC.

Statistics	THR	BNS
BIAS	1.54	1.4
POD	0.58	0.49
FAR	0.63	0.65
Log OR	1.34	1.05
ETS	0.15	0.12
HK	0.31	0.24

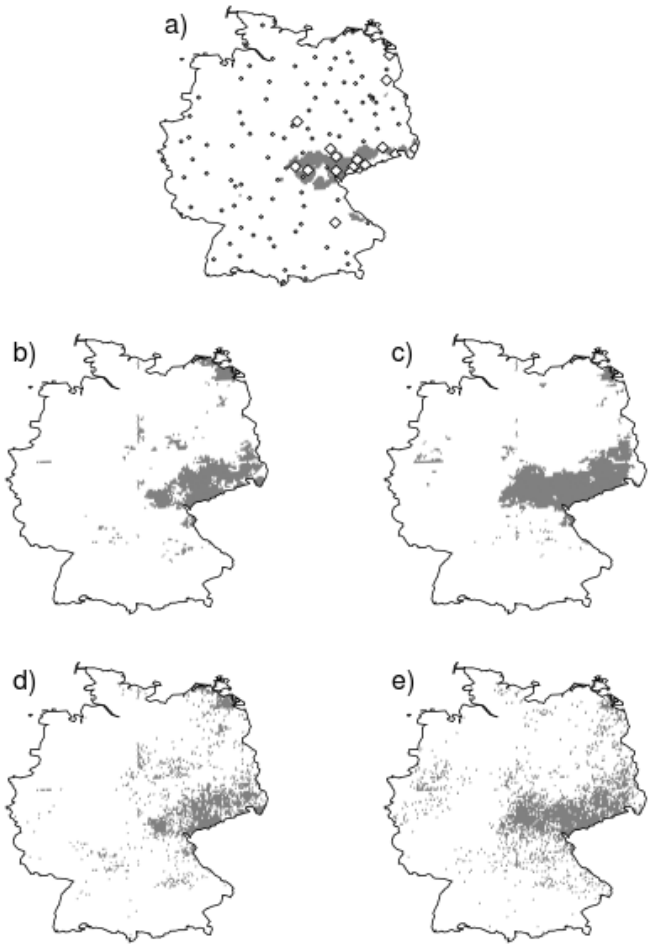

Figure E.3.: Same as Fig. 9.5 but for the precipitation case on December 14th, 2010 at 13 UTC.

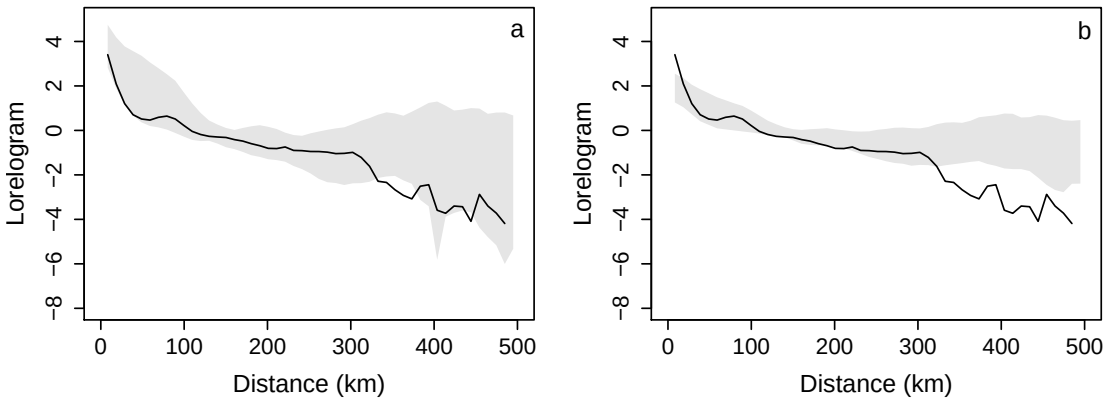

Figure E.4.: Same as Fig. 9.6, but for the precipitation case on December 14th, 2010 at 13 UTC.

For the precipitation case in category C, namely on June 05th, 2011 at 16 UTC, radar and training stations are in good agreement (Figure E.5 a)). There is also a significant correlation of -0.24 between precipitation and IRBT, as shown in chapter 8.

Precipitation ≥ 0.1 mm/h observed by radar is underestimated by HIRAIN (BIAS < 1) (table E.3). Only 29% (THR) and 28% (BNS) of the observed occurrences are predicted by HIRAIN and on the costs of a FAR of 33% and 35%, respectively. This is reinforced by the fields of $\boldsymbol{y}^p$ in Figure E.5. The odds of making a hit is greater than the odds of making a false alarm in both methods (log odds ratio > 0). There is low skill regarding the correct predictions accounting for hits to happen by chance (ETS of 0.07 (THR) and 0.05 (BNS)). The ability to separate between occurrence and non-occurrence of precipitation ≥ 0.1 mm/h is given by the HK of 0.13 (THR) and 0.10 (BNS).

Although there is not much difference between the categorical statistics for THR and BNS (table E.3), the spatial structure obtained with THR is different from the one of BNS (Figures E.5 and E.6). The spatial structure in the realisations obtained with BNS are more similar to that one of radar, than the realisations produced by THR. In this case $P_t = 0.4$ does not seem to be the most appropriate probability to threshold the realisations of $\boldsymbol{s}^p$. BNS is more able than THR to detect precipitaiton occurrence in the area that extends from north-

Table E.3.: Categorical statistics for the precipitation case on June 05th, 2011 at 16 UTC.

Statistics	THR	BNS
BIAS	0.44	0.44
POD	0.29	0.28
FAR	0.33	0.35
Log OR	0.75	0.59
ETS	0.07	0.05
HK	0.13	0.10

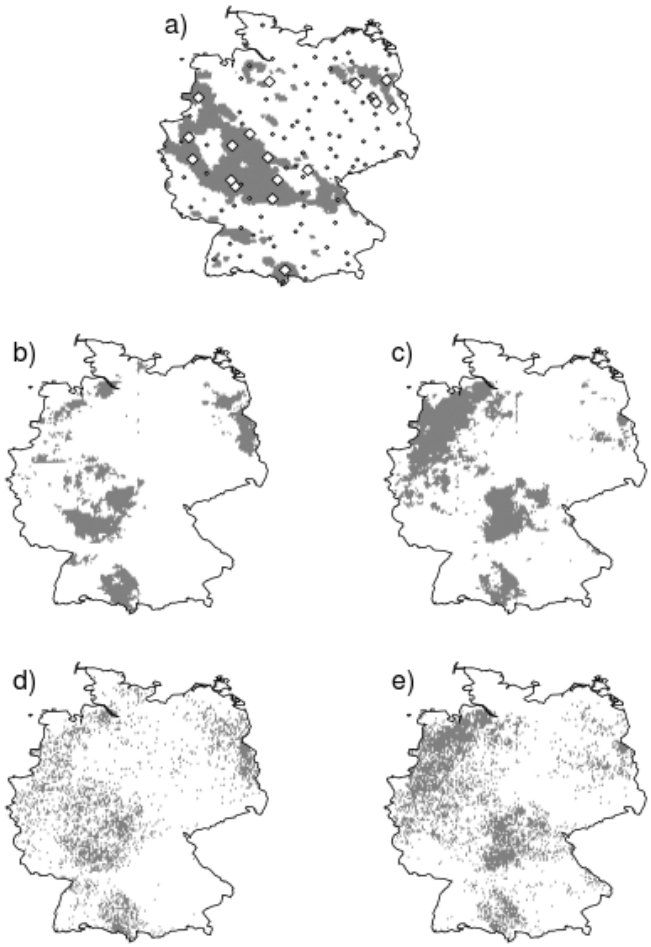

Figure E.5.: Same as Fig. 9.5 but for the precipitation case on June 05th, 2011 at 16 UTC.

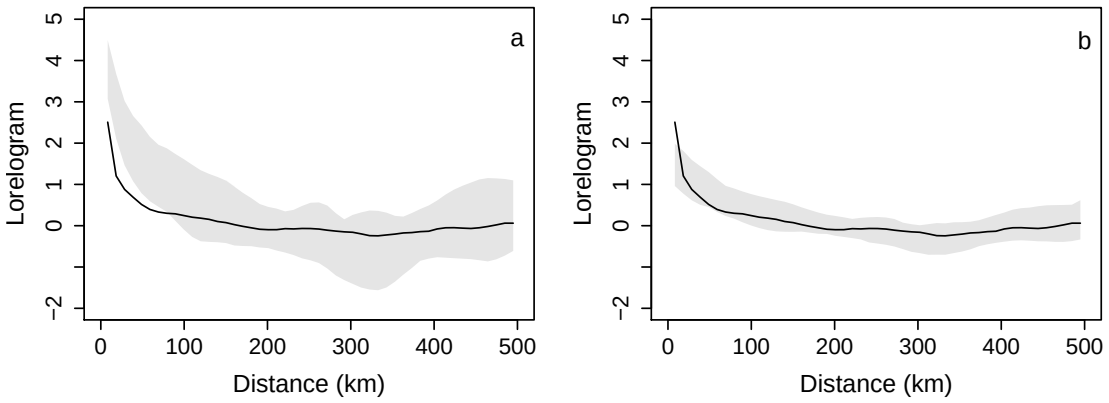

Figure E.6.: Same as Fig. 9.6, but for the precipitation case on June 05th, 2011 at 16 UTC.

western to southeastern Germany. However, BNS is not able to reproduce the neighboring spatial dependence.

These results are reinforced by the lorelograms in Figure E.6. Except for nearby locations, the radar lorelogram can be assumed to be a realisation of BNS. The THR is able to reproduce the spatial dependence of radar only at distances larger than approximately 95 km. At lower distances the log odds ratio is overestimated by this method.

Registers and Bibliography

Bibliography

Adler, R. F., Coauthors (2003). The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J.Hydrometeor.* **4**: 1147–1167.

Agrimet (Accessed 16 March 2016). RECLAMATION Managing Water in the West. URL: www.usbr.gov/pn/agrimet/.

Andersson, A., Fennig, K., Klepp, C., Bakan, S., Graßl, H., Schulz, J. (2010). The Hamburg Ocean Atmosphere Paramaters and Fluxes from Satellite Data - HOAPS-3. *Earth Syst. Sci. Data* **3**: 143–194.

Arkin, P., Ardanuy, P. (1989). Estimating Climatic-Scale Precipitation from Space: A Review. *Journal of Climate* **2**: 1229 –1238.

Arkin, P., Joyce, R., Janowiak, J. (1994). The Estimation of Global Monthly Mean Rainfall Using Infrared Satellite Data: The GOES Precipitation Index (GPI). *Remote Sensing Reviews* **11**: 107–124.

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D., Knapp, K., Cecil, L., Nelson, B., Prat, O. (2015). PERSIANN - CDR: Daily Precipitation Climate Data Records from Multisatellite Observations for Hydrological and Climate Studies. *Bull* **96**: 69–83.

Bach, L., Schraff, C., Keller, J., Hense, A. (2016). Towards a probabilistic regional reanalysis system for Europe: evaluation of precipitation from Experiments. *Tellus A* **68**: 1–21.

Bartels, H., Weigl, E., Reich, T., Lang, P., Wagner, A., Kohler, O., Gerlach, N. (2004). Projekt RADOLAN - Routineverfahren zur Online-Aneichung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer). *Deutsche Wetterdienst, Hydrometeorologie*, 1–111.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., Ziese, M. (2013). A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with samples applica-

tions including centennial (trend) analysis from 1901- present. *Earth Syst. Sci. Data* **5**: 71–99.

Bellerby, T. J., Sun, J. (2005). Probabilistic and Ensemble Representations of the Uncertainty in a IR/Microwave Satellite Precipitation Product. *J. Hydrometeor.* **6**: 1032–1044.

Bengtsson, L., Shukla, J. (1988). Integration of space and in situ observations to study global climate change. *Bulletin of the American Meteorological Society* **69**:10, 1130–1143.

Bernd, C., Rabiei, E., Haberlandt, U. (2014). Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios. *J. Hydrology* **508**: 88–101.

Berrocal, V., Raftery, A., Gneiting, T. (2008). Probabilistic Quantitative Precipitation Field Forecasting Using a Two-Stage Spatial Model. *The Annals of Applied Statistics* **2, No.4**: 1170–1193.

Bollmeyer, C., Bentzien, S., Crewell, S., Friederichs, P., Hartung, K., Hense, A., Keller, J., Keune, J., Kneifel, S., Ohlwein, C., Pscheidt, I., Redl, S., Steinke, S. (2014). Towards a high-resolution regional reanalysis for the European CORDEX domain. *Q. J. R. Meteorol. Soc.* **141**: 1–15.

Bollmeyer, C. (2015). "A High- Resolution Regional REanalysis for Europe and German - Creation and Verification with a Special Focus on the Moisture Budget." PhD thesis. Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friederich-Wilhelms-Universität Bonn.

Brier, G. (1950). Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.* **78**: 1–3.

Bromwich, D., Kuo, Y.-H., Serreze, M., Walsh, J., Bai, S., Barlage, M., Hines, K., Slater, A. (2010). Arctic System Reanalysis: Call for Community Involvement. *Eos, Transactions American Geophysical Union* **91**:2, 13–14.

Calheiros, R., Zawadzki, I. (1987). Reflectivity rain-rate relationships for radar hydrology in Brazil. *J. Climate and Appl. Meteor.* **26**: 118–132.

Chen, M., Shi, W., Xie, P., Silva, V., Kousky, V., Higgins, R., Janowiak, J. (2008). Assessing objective techniques for gauge-based analyses of global daily precipitation. *J. Geophys. Res.* **113**:D04110, 1–13.

Chen, M., Xie, P., Janowiak, J., Arkin, P. (2002). Global Land Precipitation: A 50-yr Monthly Analysis Based on Gauge Observations. *J. Hydrometeor.* **3**: 249–266.

Christensen, O., Moller, J., Waagepetersen, R. (2001). Geometric ergodicity of Metropolis Hastings algorithms for conditional simulation in generalised linear mixed models. *Methodology and Computing in Applied Probability* **3**: 309–327.

Christensen, O., Ribeiro Jr., P. (2002). geoRglm: a package for generalized linear spatial models. *R News* **2**:2. ISSN 1609-3631, 26–28. URL: `http://cran.R-project.org/doc/Rnews`.

Christensen, O., Waagepetersen, R. (2002). Bayesian Prediction of Spatial Count Data Using Generalized Linear Mixed Models. *Biometrics* **58**: 280–286.

Clark, M., Slater, A. (2006). Probabilistic Quantitative Precipitation Estimation in Complex Terrain. *Journal of Hydrometeorology* **7**: 3–22.

Cressie, N., Wikle, C. (2011). *Statistics for Spatio-Temporal Data*. Wiley New Jersey.

Crosson, W., Duchon, C., Raghavan, R., Goodman, S. (1996). Assessment of Rainfall Estimates Using a Standard Z-R Relationship and the Probability Matching Method Applied to Composite Radar Data in Central Florida. *J. Appl. Meteor.* **35**: 1203–1219.

Current Satellites (Accessed 16 March 2016). European Organization for the Exploitation of Meteorological Satellites. URL: `http://www.eumetsat.int/website/home/Satellites/CurrentSatellites/Meteosat/index.html`.

Dee, D. P., Uppala, S., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A., Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hersbach, H., Hólm, E., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A., Monge-Sanz, B., Morcrette, J.-J., Park, B.-K., Peubey, C., Rosnay, P. D., Tavolato, C., Thépaut, J.-N., Vitart, F. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society Society* **137**:656, 553–597.

De Vera, A., Terra, R. (2012). Combining CMORPH and Rain Gauges Observations over the Rio Negro Basin. *Journal of Hydrometeorology* **13**: 1799–1809.

Diggle, P. (1990). *Time Series: A Biostatistical Introduction*. Oxford University Press, New York.

Diggle, P., Ribeiro Jr., P. (2007). *Model-based Geostatistics*. Springer, New York.

Ebert, E. (1998). Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications* **15**:51–64.

Ebita, A., Kobayashi, S., Ota, Y., Moriya, M., Kumabe, R., Onogi, K., Harada, Y., Yasui, S., Miyaoka, K., Takahashi K., e. a. (2011). The Japanese 55-year Reanalysis: An Interim Report. *SOLA* **7**: 149–152.

European Centre for Medium-Range Weather Forecasts (accessed 26 Februar 2016). URL: `http://www.ecmwf.int/en/research/climate-reanalysis`.

European Reanalysis and Observations for Monitoring (Accessed 15 March 2016). URL: `http://www.euro4m.eu/`.

Fournier, J. (Apr. 1999). *Reflectivity-Rainfall Rate Relationships in Operational Meteorology*. National Weather Service Weather Forecast Office - Tallahassee, FL. URL: `http://www.srh.noaa.gov/tlh/?n=research-zrpaper`.

Gaudard, M., Karson, M., Linder, E., Sinha, D. (1999). Bayesian spatial prediction. *Environmental and Ecological Statistics* **6**: 147–171.

Gelman, A., Carlin, J., H.S.Stern, Rubin, D. (2004). *Bayesian Data Analysis*. Chapman and Hall/CRC.

German, S., German, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattn. Anal. Mach. Intel.* **6**: 721–741.

Geweke, J. (1992). *Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In Bayesian Statistics 4 (ed JM Bernado, JO Berger, AP Dawid and AFM Smith)*. Clarendon Press, Oxford.

Gilks, W., Richardson, S., Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall/CRC, London.

Groisman, P., Legates, D. (1994). The Accuracy of United States Precipitation Data. *Bull. Amer. Meteor. Soc.* **75**:3, 215–227.

Haberlandt, U. (2007). Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *J. Hydrology* **332**: 144–157.

Harris, I., Jones, P., Osborn, T., Lister, D. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *Int. J. Climatol.* **34**: 623–642.

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97–109.

Haylock, M., Hofstra, N., Tank, A. K., Klok, E., Jones, P., New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950 - 2006. *J. Geophys. Res.* **113**:

Heagerty, P. J. (1995). "Multivariate Multinomial Marginal Models." PhD thesis. Johns Hopkins Univerity, Dept. of Biostatistics.

Heidelberger, P., Welch, P. (1981). A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations. *Comm. ACM* **24**: 233–245.

– (1983). Simulation run length control in the presence of an initial transient. *Opns Res.* **31**: 1109–44.

Huffman, G., Coauthors (1997). The Global Precipitation Climatology Project (GPCP) combined precipitation data set. *Bull. Amer. Meteor. Soc.* **78**: 5–20.

Huffman, G., Adler, R., Morrissey, M., Bolvin, D., Curtis, S., Joyce, R., McGavock, B., Susskind, J. (2001). Global Precipitation at one-Degree Daily Resolution from Mulstisatellite Observations. *J. Hydrometeor.* **2**: 36–50.

Janowiak, J., Xie, P. (1999). CAMS-OPI: A Global Satellite-Rain Gauge Merged Product for Real-Time Precipitation Monitoring Applications. *J. Climate* **12**: 3335–3342.

Janssen, M. A. (1993). *Atmospheric Remote Sensing by Microwave Radiometry*. John Wiley and Sons, Inc.

Jones, C. D., Macpherson, B. (1997). A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model. *Meteorological Application* **4**: 269–277.

Joseph, R., Smith, T., Sapiano, M., Ferraro, R. (2009). A New High-Resolution Satellite-Derived Precipitation DatData for Climate Studies. *J. Hydrometeor.* **10**: 925–952.

Joyce, R., Janowiak, J., Arkin, P., Xie, P. (2004). CMORPH: A Method that Produces Global Precipitation Estimates from Passive Microwave and Infrared Data at High Spatial and Temporal Resolution. *J. Hydrometeor* **5**: 487–503.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Ropelewsk, C., Wang, J., Jenne, R., Joseph, D. (1996). The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society* **77**:3, 437–471.

Krajewski, W. (1987). Cokriging Radar-Rainfall and Rain Gage Data. *J. Geophys. Research* **92**:D8, 9571–9580.

Marshall, J., Langille, R. C., Palmer, W. M. (1947). Measurement of rainfall by RADAR. *J. Meteor.* **4**: 186–192.

Marshall, J. S., Palmer, W. M. (1948). The distributions of raindrops with size. *J. Meteor.* **9**: 327–332.

Menzel, P. (Accessed 07 March 2016). Cooperative Institute for Meteorological Satellite Studies. URL: `https://cimss.ssec.wisc.edu/rss/benevento/index.html/lecture1_4jun_planck_function.ppt`.

Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P., Ebisuzaki, W., Jovic, D., Woollen, J., Rogers, E., Berbery, E., Ek, M., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y., Manikin, G., Parrish, D., Shi, W. (2006). North American Regional Reanalysis. *Bull. Am. Meteor. Soc.* **87**:3, 343–360.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E. (1953). Equations of state calculations by fast computing machine. *J. Chem. Phys* **21**: 1087–1091.

NOAA's Geostationary and Polar-Orbiting Weather Satellites (Accessed 8 March 2016). NOAA Satellite Information System (NOAASIS). URL: `http://noaasis.noaa.gov/NOAASIS/ml/genlsatl.html`.

Onogi, K., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi K., e. a. (2007). The JRA-25 Reanalysis. *Journal of the Meteorological Society of Japan, Ser. II* **85**:3, 369–432.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S., Takacs, L., al., K. K. e. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of Climate* **24**:14, 3624–3648.

Rinehart, R. (1997). *Radar for Meteorologists*. Third Edition. University of North Dakota.

Roberts, N., Lean, H. (2008). Scale-selective verifications of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Review* **136**:78–97.

Roebeling, R. A., Holleman, I. (2009). SEVIRI rainfall retrieval and validation using weather radar observations. *J. Geophys. Res.*. **114**: D21202.

Saha, S., Moorthi, S., Pan, H.-L., Wu, X., Wang, J., Nadiga, S., Tripp, P., Kistler, R., Woollen, J., Behringer D., e. a. (2010). The NCEP climate forecast system reanalysis. *Bull. Amer. Meteor. Soc.* **91**:8, 1015–1057.

Sanso, B., Guenni, L. (1999). Venezuelan rainfall data analysed by using a Bayesian space-time model. *Appl. Statist* **48, Part 3**: 345 –362.

Sapiano, M., Arkin, P. (2009). An Intercomparison and Validation of High-Resolution Satellite Precipitation Estimates with 3-Hourly Gauge Data. *J. Hydrometeor.* **10**: 149–166.

Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., Ratier, A. (2002). An Introduction to Meteosat Second Generation (MSG). *Amer. Meteor. Soc.* 977–991.

Schmid, J. (Accessed 08 March 2016). *The SEVIRI Instrument*. European Organization for the Exploitation of Meteorological Satellites (EUMETSAT). URL: `http://www.eumetsat.int/website/wcm/idc/groups/ops/documents/document/mday/mde1/~edisp/pdf_ten_msg_seviri_instrument.pdf`.

Schwartz, C., Kain, J., Weiss, S., Xue, M., Bright, D., Kong, F., Thomas, K., Levit, J., Coniglio, M., Wandishin, M. (2010). Toward Improved Convection-Allowing Ensembles: Model Physics Sensitivities and Optimizing Probabilistic Guidance with Small Ensemble Membership. *Weather and Forecasting* **25**:263–280.

Simmer, C., Adrian, G., Jones, S., Wirth, V., Goeber, M., Hohenegger, C., Janjic, T., Keller, J., Ohlwein, C., Seifert, A., Troemel, S., Ulbrich, T., Wapler, K., Weissmann, M., Keller, J., Masbou, M., Meilinger, S., Riss, N., Schomburg, A., Stein, C., Vormann, A., Weingärtn, C. (2015). HErZ - The German Hans-Ertel Centre for Weather Research. *Bulletin of the American Meteorological Society*.

Stauffer, D., Seaman, N. (1990). Use of four dimensional data assimilation in an limited-area mesoscale model. Part I: Experiments with synoptic-scale data. *Monthly Weather Review* **118**: 1315–1326.

Stephan, K., Klink, S., Schraff, C. (2008). Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD. *Q.J.R. Meteorol. Soc.* **134**: 1315–1326.

Stephenson, D. (2000). Use of the "Odds Ratio" for Diagnosing Forecast Skill. *Weather and Forecasting* **15**: 221–232.

Strangeways, I. (2007). *PRECIPITATION Theory, Measurement and Distribution*. Cambridge University Press.

Trenberth, K., Olson, J. (1988). An evaluation and intercomparison of global analyses from the National Meteorological Center and the European Centre for Medium-Range Weather Forecasts. *Bull. Amer. Meteor. Soc.* **69**: 1047 –1057.

Uppala, S. M., Kållberg, P., Simmons, A., Andrae, U., Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R., Andersson, E., Arpe, K., Balmaseda, M., Beljaars, A. C. M., Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., Mcnally, A., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N., Saunders, R., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., Woollen, J. (2005). The ERA-40 reanalysis. *Quarterly Journal of the Royal Meteorological Society* **131**:612, 2961 –3012.

Vila, D. A., Goncalvez, L. G. G. d., Toll, D. L., Rozante, J. R. (2009). Statistical evaluation of combined daily gauge observations and rainfall satellite estimates over continental South America. *America. J. Hydrometeor.* **10**: 533–543.

Villarini, G., Krajewski, W., Ciach, G., Zimmerman, D. (2009). Product-error-driven generator of probable rainfall conditioned on WSR-88D precipitation estimates. *Water Resources Research* **45**: 1–11.

Vischel, T., Lebel, T., Massuel, S., Cappelaere, B. (2009). Conditional simulation schemes of rain fields and their application to rainfall-runoff modeling studies in the Sahel. *Journal of Hydrology* **375**: 273–286.

Waagepetersen, R., Ibanez-Escriche, N., Sorensen, D. (2008). A comparison of strategies for Markov chain Monte Carlo computation in quantitative genetics. *Genet. Sel. Evol.* **40**: 161–176.

Wahl, S., Bollmeyer, C., Crewell, S., Figura, C., Friederichs, P., Hense, A., Keller, J., Ohlwein, C. (2017). A novel convective-scale regional reanalysis COSMO-REA2: Improving the representation of precipitation. *Meteorologische Zeitschrift*.

Xiaolan, L. W., Lin, A. (2015). An Algorithm for integrating satellite precipitation estimates with in situ precipitation data on a pentad time scale. *J. Geophys. Res. Atmos* **120**: 1–17.

Xie, P., Arkin, P. (1997). Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs. *Bull. Amer. Meteor. Soc.* **78**:11, 2539–2558.

Xie, P., Janowiak, J. E., Arkin, P. A., Adler, R., Gruber, A., Ferraro, R., Huffman, G. J., Curtis, S. (2003). GPCP pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. *Journal of Climate* **16**: 2197–2214.

Yang, G., Slingo, J. (2001). The diurnal cycle in the Tropics. *Mon. Wea. Rev* **129**: 784–801.

Yilmaz, M., Houser, P., Shrestha, R., Anantharaj, V. (2010). Optimally Merging Precipitation to Minimize Land Surface Modeling Errors. *J. Appl. Meteor. and Climat.* **49**:3, 415–423.

BONNER METEOROLOGISCHE ABHANDLUNGEN

Herausgegeben vom Meteorologischen Institut der Universität Bonn durch Prof. Dr. H. FLOHN (Hefte 1-25), Prof. Dr. M. HANTEL (Hefte 26-35), Prof. Dr. H.-D. SCHILLING (Hefte 36-39), Prof. Dr. H. KRAUS (Hefte 40-49), ab Heft 50 durch Prof. Dr. A. HENSE.

Heft 1-63: siehe `http://www.meteo.uni-bonn.de/bibliothek/bma`

64-77: open access, verfügbar unter `https://uni-bn.de/kpSDaQffe1`

Heft 64: ***Michael Weniger***: Stochastic parameterization: a rigorous approach to stochastic three-dimensional primitive equations, 2014, 148 S. + XV.

Heft 65: ***Andreas Röpnack***: Bayesian model verification: predictability of convective conditions based on EPS forecasts and observations, 2014, 152 S. + VI.

Heft 66: ***Thorsten Simon***: Statistical and Dynamical Downscaling of Numerical Climate Simulations: Enhancement and Evaluation for East Asia, 2014, 48 S. + VII. + Anhänge

Heft 67: ***Elham Rahmani***: The Effect of Climate Change on Wheat in Iran, 2014, [erschienen] 2015, 96 S. + XIII.

Heft 68: ***Pablo A. Saavedra Garfias***: Retrieval of Cloud and Rainwater from Ground-Based Passive Microwave Observations with the Multi-frequency Dual-polarized Radiometer ADMIRARI, 2014, [erschienen] 2015, 168 S. + XIII.

Heft 69: ***Christoph Bollmeyer***: A high-resolution regional reanalysis for Europe and Germany - Creation and Verification with a special focus on the moisture budget, 2015, 103 S. + IX.

Heft 70: ***A S M Mostaquimur Rahman***: Influence of subsurface hydrodynamics on the lower atmosphere at the catchment scale, 2015, 98 S. + XVI.

Heft 71: ***Sabrina Wahl***: Uncertainty in mesoscale numerical weather prediction: probabilistic forecasting of precipitation, 2015, 108 S.

Heft 72: ***Markus Übel***: Simulation of mesoscale patterns and diurnal variations of atmospheric CO_2 mixing ratios with the model system TerrSysMP-CO_2, 2015, [erschienen] 2016, 158 S. + II

Heft 73: ***Christian Bernardus Maria Weijenborg***: Characteristics of Potential Vorticity anomalies associated with mesoscale extremes in the extratropical troposphere, 2015, [erschienen] 2016, 151 S. + XI

Heft 74: ***Muhammad Kaleem***: A sensitivity study of decadal climate prediction to aerosol variability using ECHAM6-HAM (GCM), 2016, 98 S. + XII

Heft 75: ***Theresa Bick***: 3D Radar reflectivity assimilation with an ensemble Kalman filter on the convective scale, 2016, [erschienen] 2017, 96 S. + IX

Heft 76: ***Zied Ben Bouallegue***: Verification and post-processing of ensemble weather forecasts for renewable energy applications, 2017, 119 S.

Heft 77: ***Julia Lutz***: Improvements and application of the STatistical Analogue Resampling Scheme STARS, 2016, [erschienen] 2017, 103 S.

Heft 78: ***Benno Michael Thoma***: Palaeoclimate Reconstruction in the Levant and on the Balkans, 2016, [erschienen] 2017, XVI, 266 S.

Heft 79: ***Ieda Pscheidt***: Generating high resolution precipitation conditional on rainfall observations and satellite data, 2017, V, 173 S.

METEOROLOGISCHES INSTITUT
MATHEMATISCH NATURWISSENSCHAFTLICHE FAKULTÄT
UNIVERSITÄT BONN

Meteorologisches Institut
Universität Bonn