

A COMPARISON OF SPACE-TIME MODELING APPROACHES FOR EXTREME PRECIPITATION IN GERMANY BETWEEN 1996 AND 2016

A thesis paper for the degree of Bachelor of Arts (B.A.)

Wintersemester 2022

By

Ekin Gülhan

Supervisor: Dr. Isa Marques
Enrolment number: 21675680
e-mail: ekin.guelhan(at)stud.uni-goettingen.de
Address: Hannoversche Straße 134, 37077 Göttingen
Date: 2023-03-14

Contents

Preface	I
1 Introduction	1
Data	5
2 Framing the prediction of extreme precipitation as a binary classification of precipitation events	6
2.1 Defining precipitation events and extreme precipitation events	7
2.2 A standard of assessing the truth of a binary classification of precipitation events	8
2.3 Modelling a binary classifier model: The concept of classification logic	10
2.4 Defining a classification logic based on deterministic and statistical models	11
3 Appropriate choices for the classification logic	14
3.1 Principles that govern patterns of precipitation	14
3.2 Spatio-Temporal Autocorrelation of precipitation heights	15
3.3 IDW based interpolation as an appropriate choice for the classification logic	19
3.4 A trained GAMM as an appropriate choice for the classification logic	20
3.4.1 Classical Linear Regression	20
3.4.2 Logistic Regression Models as <i>Generalized Linear Models</i>	21
3.4.3 Generalized Additive Mixed Model with a logit-link function	22
4 A competitive evaluation of the performance of the classifier models	24
4.1 How confusion-matrix based performance indicators quantify classifier performance	25
4.2 Confusion-matrix and performance indicators for the IDW based classifier	27
4.3 Consufion-matrix and performance indicators for the GAMM based classifier	27

4.4 A comparison of the performance indicators of both classifiers	27
5 Conclusion	27
Appendix	III
Defining the classification logic	III
Classical Linear Regression	III
5.0.1 Logistic Regression	VI
5.0.2 Generalized Additive Mixed Model	VIII
Inverse Distance Weighting Based Interpolation	IX
Measurements of spatial, temporal and spatio-temporal autocorrelation	XI
Haversine Formula	XI
Empirical Spatio-Temporal Covariance	XIII
Empirical Spatial Covariance	XVI
Empirical Temporal Covariance	XVIII
Empirical Spatial Semivariogram	XX
Empirical Temporal Semivariogram	XXII
6 References	XXIV

Preface

Dear reader,

the author of this thesis paid special attention the creation of a reproducible and dynamic reasearch paper. This .pdf version of the thesis has been created from a rmarkdown file¹ `BachelorThesis.rmd` and this .rmd file is available in a GitHub repository². It is available alongside all the R code written for different analytical purposes and which is encapsulated in different thematically named R scripts. The repository on GitHub has been ordered in a way that allows easy and intuitive access and navigation: Input and Output, source code and the .rmd file are divided in sepearate directories, which are then subdivided based on format of the data and the topic of the content. Whenever code has been written, great attention has been paid to readability. To allow for such readability, tidyverse syntax and functions have been used consistently. This implementation of the tidyverse does also reflect the authors understanding of an effective way to account for the data-flow centered nature of the programming language R for the merit of data science.

As a project, the thesis has been made to evolve around a file of type **R Project** to further enable reproducibility and mobility. The `BachelorThesis.R Project` file can be considered as the ultimate reference point within the toplevel of the project directory. When browsing through the directory or trying to reproduce a specific analytical step or even the whole thesis analysis, the reader should regard `BachelorThesis.R Project` as the entry point to be opened as an unconditional first step. Such restriction is neccessary, as it allows for the mobility of the thesis in the first place, when used in combination with the R packages `here` and `renv`. In this context, mobility is understood as a characteristic of the most top-level directory of the thesis, to be the sole reference point for the .rmd file and all R scripts. When knitting `BachelorThesis.rmd` to `BachelorThesis.pdf` or when executing any R script, there is no need to inconvenience oneself with paying attention to setting working directories and defining correct absolute paths.

While the reader can forego inconvenient absolute paths with the help of `BachelorThesis.R Project` and `here`, the combination of `BachelorThesis.R Project` and the R package `renv`

¹Note that this .pdf file was successfully knitted from the .rmd file on a windows PC with 8 gigabytes of RAM without any other programs running except for system related services.

²Repository link: <https://github.com/ekingn/BachelorThesis>.
Access to the repository on request. Reach out to me via ekin.guelhan@posteo.de

allows for a secondary project-specific library. A `renv.lock` file snapshots the packages in the respective versions, in which they were used within the thesis project, in a descriptive list. By utilizing the `renv.lock` file through the package `renv`, any reader can easily install all the packages in the exact versions required for the reproduction of analytical steps and creating the present `BachelorThesis.pdf` file. The packages are assigned to a secondary project library, such that the primary user library remains unchanged.

In summary, if the readers intent goes beyond reading the `BachelorThesis.pdf`, he or she should do the following:

1. Open the file `BachelorThesis.R Project`
2. If not already done, install and require the R package `here` and `renv`
3. Execute `renv::restore()`.
4. Possibly repeat the last step multiple times.

A secondary project-specific library is available and used once these steps are complete and the thesis can be reproduced without effort.

The author also wants to point out, that that the contents of the project directory go beyond what is explicitly included in the `BachelorThesis.pdf` file. A plethora of data exploration results including various data visualizations regarding spatial and temporal means or Hovmöller diagrams, an animation of the variation of precipitation heights against the variation of spatial location and time, as well as others more, can be found in the GitHub repository as well. I therefore invite the curious reader to browse its content in the hope, that he or she finds something of interest.

When reading equations and formulae, the reader should look out for symbols and strings in bold-font, which denote vectors.

I finally want to thank Dr. Isa Marques for her through-out available council and insightful as well as helpful feedback.

Ekin Gülhan

2023-03-14

1 Introduction

Extreme precipitation, though rare, may result in floodings of a magnitude similar to those that occurred 2002 and 2013 and affected Europe and Germany in particular, leading to loss of life and property (cf. Kreibich et al. [2017], p. 2075)(cf. Kuhlicke et al. [2016], p. 1).

Humans may not prevent the occurrence of such detrimental natural incidents altogether, but they can and have installed a system of warnings and countermeasures to mitigate the damage as best as they can. When extreme precipitation affected Germany in 2002 and 2013 respectively, the country already had its specific flooding-damage mitigation system in work. An essential component of such a system is an effective flood warning system. Two parts of any such warning systems are

"the detection of potentially hazardous situations" and

"rules on when, how and whom to warn in case of rising flood water levels and what to communicate in order to activate organization in charge of civil protection".

(cf. Kreibich et al. [2017], p. 2078) Especially in the context of the forwarding of flood-warnings, research was conducted in the aftermath of the floodings of 2002, 2013 and 2021 in the form of surveys of the German residents in affected areas (cf. Kreibich et al. [2017])(cf. Thielen et al. [2022]). They reveal, that a substantial group of the recipients of such warnings did not know how to operationalize the what they received. In other words: These citizens did not know what to do. Presently, there is no specific research as to the reasons for this lack of decisiveness or decision paralysis respectively. It appears hence, that the warnings that were forwarded³ and in most cases indeed received by the affected residents (at least during the flooding of 2013 (cf. Kreibich et al. [2017], p.2075)), did not elicit the expected response.

Surveys of those affected residents however, who had already experienced a flooding before, show a significant reduction in the aforementioned lack of decisiveness compared to those residents, who had experienced a flooding for the first time. This is likely the result of learned behaviour, to which the flooding served as catalyzer.

Effective flooding management can obviously not rely on German residents to survive a flooding

³In Germany, the federal states are primarily responsible for flood management (cf. @Kuhlicke.2016, p. 2).

once to improve the effectiveness of flood-management. On this basis, the present thesis paper asserts, that research into the transformation of existent flood-warning systems may be warranted. It furthermore asserts, that such research could focus on the implementation of a categorical warning typology; if the residents get a clear and unambiguous instruction, decision paralysis may be prevented⁴.

The idea of a categorical warning typology refers to a system, where distinct warnings are made up of two-components:

An assesement of the level of flooding severity for a given spatio-temporal location. Flooding severity is to be measured on a categorical scale.

A clear and unambiguous instruction for the residents, that is distinct for each assessed categorical level of flooding severity.

This thesis contributes to a small area within this research field. *It is specifically contributes research to the ways of defining and measuring the level of dichotomously conceptualized flooding severity on the basis of predictions of extreme precipitation.* A simple sketch of a dichotomous concept of flooding severity on the basis of the predicted binary precipitation status and the forwarding of associated distinct warning and action prompts, reads as follows:

Category of predicted flooding severity	Conditions
C_0	$\widehat{EPE} = 0$ (non-extreme Precipitation)
C_1	$\widehat{EPE} = 1$ (extreme Precipitation)

Table 1: A dichotomous typology of flooding severity

and

$$f(C_i) \begin{cases} \text{No action} & \text{if } i = 0 \\ \text{Follow evacuation orders and seek higher ground immediately} & \text{if } i = 1 \end{cases} \quad (1.1)$$

$$(1.2)$$

⁴At the current stage, this is but a working, hypothesis that requires efforts of social scientists towards validation or falsification

where

$f()$ is a function that maps the assessed flooding category C_i to the action, that residents are advised to perform.

C_i is the label or category of flooding severity assessed on the basis of the predicted status of precipitation (extreme/non-extreme)

\widehat{EPE} is a binary prediction of the status of the precipitation (extreme/non-extreme) at a spatio-temporal location of interest.

Following this concept, a level of flooding severity from a dichotomous scala is assigned to a spatio-temporal location on the basis of the predicted status of precipitation at that location. For a spatio-temporal location, an flood-warning forwarding algorithm would hence generate a prediction of the status of precipitation, on whose basis a label of flooding severity is assigned. This label or level of flooding severity would be translated into one of two warnings and actions prompts, which are then forwarded to the residents at spatio-temporal location of interest. Within this context of research on enhancing the German flood-management system and more precisely on the prediction of extreme precipitation for the assignment of dichotomous levels flooding severity, the research question of this thesis paper reads as follows:

Research Question

How do two classifier models compare performance-wise in predicting the occurrence of extreme precipitation in Germany between 1996 and 2016, if

the classification logic in one classifier is defined on the basis of inverse distance weighting (IDW) based interpolation

the classification logic in the other classifier is defined on the basis of a generalized additive mixed model (GAMM) with a logit link?

An answer to this question is provided in three steps and these built the paper's main body. The first step is an act of framing. The prediction of the precipitation of the binary status of prediction as either extreme or non-extreme is framed as the binary classification of precipitation events as

either extreme or non-extreme precipitation events. Precipitation and extreme precipitation events are defined first and afterwards the process of modelling a binary classifier model is described while focusing on the importance and definition of a classification logic.

The second step is a consideration of appropriate choices for the definition of the classification logic. The exposition of appropriate choices is developed on the basis of their fit to the principles, that are usually asserted to govern the phenomenon precipitation. Here, IDW based interpolation and a trained GAMM are both argued as appropriate alternatives in defining the classification logic.

In the final third step, a competitive evaluation of the classifiers is conducted with reference to their performance in the prediction of extreme precipitation.

Data

As a preliminary to this main body however, a short introduction to the analyzed data is due.

The training data used in this thesis is meteorological data sampled by 438 German weather stations whose spatial locations are depicted on figure 1.

It has a panel format and is irregular, in the sense that the number of observations are not the same for all 438 weather stations. The years of observation range between 1996 and 2012. This training data is complemented by a second data set of meteorological data, the hold-out data. The years, in which the meteorological observations in the hold-out data were made, range between 2013 and 2016. This split of the data into a training-data subset and a hold-out data subset is a common data-scientific practice (cf. Zumel and Mount [2014]). The two classifier models are built (trained) on the basis of the above mentioned training data, for whom a small representative slice is presented in table 2 below. Afterwards the hold-out data is utilized as input to the trained model and the correspondingly generated classifications of precipitation evaluated with reference to model performance. These names for the two subsets of the analyzed data reflect

the purpose of the corresponding data split: The training data is the basis for the estimation procedure, during which the classifier models *learn* a specific classification logic. Data scientists however regard this training data set as only partially appropriate to assess the performance of a built / trained classifier model out of concern for bias. The concern for bias is essentially related to the problem of overfitting. Overfitting occurs, when the classifier model adopts the variation in the training data too closely during the training procedure. This might result in poor performance once the trained model is given different data as input. Such will be the case, if the classification logic is only valid for the data which the classifier model was trained upon, and not for the population

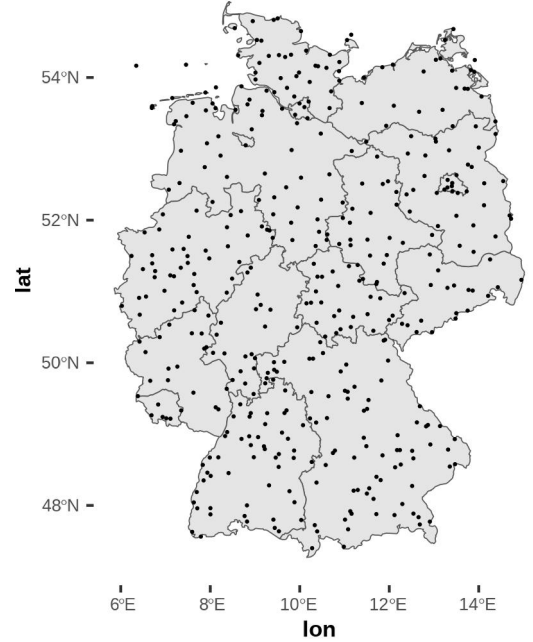


Figure 1: Map of 438 German weather stations in the training data. Own representation.

data. The hold-out data is a dataset hence used instead of the training training data for model evaluation purposes; it is analyzed, how well the built models are “holding out” the data. This concludes a short introduction to the analyzed data. The following sections present the three aforementioned steps.

name	index	lon	lat	year	month	day	season	days_since_earliest_observation	rain	alt	mean_temperature	monthly_mean_precipitation	EPE
Aachen-Orsbach	1	6.0244	50.7983	2011	4	1	Spring	5569	0.0	230.99	13.3	41.15	0
Aachen-Orsbach	1	6.0244	50.7983	2011	4	2	Spring	5570	0.0	230.99	16.8	41.15	0
Aachen-Orsbach	1	6.0244	50.7983	2011	4	3	Spring	5571	9.9	230.99	10.8	41.15	0
Aachen-Orsbach	1	6.0244	50.7983	2011	4	4	Spring	5572	0.0	230.99	9.2	41.15	0
Aachen-Orsbach	1	6.0244	50.7983	2011	4	5	Spring	5573	0.9	230.99	8.8	41.15	0

¹ 'index' (=def.) indexnumber of the weather station

² 'mean_temperature' (=def.) daily mean temperature

³ 'rain' (=def.) observed precipitation height in millimetre (mm)

⁴ 'year', 'month' and 'day' (=def.) the year, month and day of an observation

⁵ 'season' (=def.) the corresponding one among the annual seasons

⁶ 'days_since_earliest_observation' (=def.) a type of julian date. The number of days, that have passed since the earliest observation in the whole training data.

⁷ 'alt' (=def.) altitude level of weather station

⁸ 'lon' and 'lat' (=def.) longitudinal and latitudinal degree of weather station

⁹ 'name' (=def.) the city associated with the weather station

¹⁰ 'EPE' (=def.) binary variable that indicates the true classification of the respective precipitation event. 1 := Extreme Precipitation Event. 0 := Non-Extreme Precipitation Event

¹¹ 'monthly_mean_precipitation' (=def.) typical total precipitation of the respective month for the respective weather station

Figure 2: Rows 1-5 of the training data. Own representation

2 Framing the prediction of extreme precipitation as a binary classification of precipitation events

The research question asks for a competitive evaluation of the performance of the classifier models in predicting extreme precipitation. The problem of modelling a binary response such as the occurrence of extreme precipitation at a given spatio-temporal location is indeed generally framed as a binary classification problem within (data) scientific literature(cf. Zumel and Mount [2014])(cf. Hosmer et al. [2013], p.18). In the next section, such framing will be conducted explicitly before introducing the two classifier models in detail.

Let Extreme Precipitation Event (EPE) be the class of interest and let precipitation events be the subjects, for whom the membership inside this class is to be modelled: Let a precipitation event $PE_{s_i,t}$ have been observed at the spatio-temporal location $s_i, t \in ST$, where s_i denotes the spatial location of the observation and t denotes the specific day of the observation, a model is to be trained. And let ST denote the spatio-temporal domain of the training data

set. Then a model is to be trained, that either $PE_{s_i,t}$ as an EPE or as an non-EPE. In other words: It is to classify $PE_{s_i,t}$ as either EPE or non-EPE.

2.1 Defining precipitation events and extreme precipitation events

The precipitation events referred to in this context are non-empirical, that is to say, they are not directly observable. They are a construct and a generated classification of a precipitation event as an extreme precipitation event is at best correctly associated with the warning and action prompt, that is forwarded on its basis. The definition of precipitation events generally involve either one or a combination of three usually measured empirical characteristics of precipitation, namely precipitation height (PH), duration and intensity of precipitation (cf. Trömel and Schönwiese [2007])(cf. Keller and Atzl [2014])(cf. Thielen et al. [2016])(cf. Kreibich et al. [2017])(cf. Wagenaar et al. [2018])(cf. Gimeno et al. [2022]). For the purpose of this thesis, precipitation events are defined on the basis of PH only. There is in fact a great variation of its spatio-temporal distribution and the most common observation is, that there is no precipitation at all, or $PH_{s_i,t} = 0$. Although rare, PH can realize in extreme forms. When it does, it is likely to our detriment.

The definition of extreme precipitation events naturally follows this logic and involves precipitation height as well. Some researchers take the approach to define EPE on the basis of the lower or upper tail of the space-time distribution of precipitation height, which is to say, on the basis of a specific quantile (cf. Trömel and Schönwiese [2007])(cf. Gimeno et al. [2022]). A $PE_{s_i,t}$ would be considered an EPE then, if $PH_{s_i,t}$ exceeded the respective α -quantile value. Whether PE that have been classified as EPE on the basis of such an approach would however warrant any action at all, is highly questionable: If extreme precipitation events are for example defined utilizing the 95th percentile as threshold value, which is the approach in Trömel (2007), then the ratio of extreme precipitation events to all precipitation events would be indeed 5%; the absolute number of extreme precipitation events however were 86392. It contradicts common sense to assert, that the federal states of Germany had to warn their citizens in total of 86392 cases between 1996 and 2012, because a extreme precipitation event is occurring that might cause a dangerous flooding. This thesis took an alternative approach. The calculation of the threshold value and the subsequent classification of precipitation events as extreme read as follows:

First, let $PH_{s,y,m,d}$ be the daily total precipitation height observed at spatial location s of a German weather station and y , m and d denote a year, month and day of the month (compared to previous notation, $y, m, d := t$ and $s_i := s$). Also, let $PH_{s,y,m}$ be the total precipitation height of the month m in the year y at location s .

This monthly total precipitation height $PH_{s,y,m}$ then is calculated as:

$$PH_{s,y,m} = \sum_{d \in m} PH_{s,y,m,d} \quad (2.1)$$

For the same spatial location s , the average of all monthly total precipitation heights for the month m observed in the different years in the training data y in $Y_{training} = \{1996, 1997, \dots, 2012\}$ can be calculated as:

$$\overline{PH}_{s,m} = \frac{1}{N_{s,m}} \sum_{y \in Y} PH_{s,y,m} \quad (2.2)$$

where $N_{s,m}$ is the number of years y , for whom precipitation height has been observed in the month m at the location s .

The equation 2.2 calculates the average monthly total precipitation height of month m at location s across all years y . It represents the typical total precipitation height for that month at that location.⁵ The logic of classification of precipitation events as extreme or non-extreme finally utilizes this typical monthly precipitation height as

$$PE_{s_i,t} \text{ is classified as EPE, if } PH_{s,y,m,d} > \overline{PH}_{s,m} \quad (2.3)$$

$$PE_{s_i,t} \text{ is classified as non-EPE, if } \overline{PH}_{s,m} > PH_{s,y,m,d} \quad (2.4)$$

where $\overline{PH}_{s,m}$ serves as the threshold / cut-off value.

2.2 A standard of assessing the truth of a binary classification of precipitation events

In the context of the scientific motivation for this paper, it was earlier mentioned, that two classifier models are to be trained to generate classifications. And that flooding severity labels / levels are

⁵Note that the monthly typical total precipitation height is very likely unique for each spatial location.

assigned to a spatio-temporal location on the basis of the generated classifications, such that ultimately a distinct warning and action prompt is forwarded to the residents at that location. Against this background, the role of this particular classification logic is to provide a so-called Standard of Truth (S.o.T.)⁶. A S.o.T. establishes an unconditional and unambiguous reference for assessing the validity or correctness of a generated classification. This means: If the generated classification aligns with the classification logic described above (the S.o.T.), then it is labeled as *true* / *correct*. Another way to express the necessity for a S.o.T. is, that it defines the classifications, that the researcher *wants* the trained classifiers to generate. In the context of this thesis, this could be because classifications that align with the S.o.T. have proved to significantly correlated with either non-dangerous or dangerous amounts of precipitation in past research⁷. The easiest way for using the S.o.T. as a reference of *truth* practically is to determine a set of classifications who align with the S.o.T. for $PE_{s_i,t} \forall i, t \in ST$. These *true classifications* are best procured in the very beginning of the analysis and each respective classification generated by a trained classifier for $PE_{s_i,t}$ can then later on be compared with the corresponding true classification for the same $PE_{s_i,t}$. Such a set of true classifications can be constructed as a binary variable in the training data and the process of the construction can be expressed as follows:

Let $EPE_{s_i,t}$ thus denote the classification of $PE_{s_i,t}$ which aligns with the above mentioned S.o.T..

Then

$$EPE_{s_i,t} = EPE_{s,y,m,d} = \begin{cases} 1, & \text{if } PH_{s,y,m,d} > \overline{PH}_{s,m} \\ 0, & \text{if } \overline{PH}_{s,m} > PH_{s,y,m,d} \end{cases} \quad (2.5)$$

⁶It is necessary to establish a S.o.T., as model evaluation were impossible, if a specific classification logic were not labeled unconditionally *true* at some point. It may not be obvious, but a classification logic defined on the basis of IDW based interpolation or a trained GAMM do not necessarily generate classifications, that are *true* by the S.o.T..

⁷It is an important part of future research, to establish, whether extreme precipitation events as they are defined in this chapter indeed warrant any action. Given the limitations to this thesis however, it shall be assumed, that this is the case, notwithstanding the relevance of a critical analysis of this assumption. Compared to aforementioned approach of the 95th percentile as threshold-value, the S.o.T. utilized in this thesis at generates a significantly smaller number extreme precipitation events (102)

where

EPE is a binary variable that indicates the true classifications of precipitation events as either extreme or non-extreme

$EPE_{s_i,t}$ ($EPE_{s,y,m,d}$) is the value of that variable for $PE_{s_i,t}$

$EPE_{s_i,t}$ ($EPE_{s,y,m,d}$) evaluates to 1, if the true classification of $PE_{s_i,t}$ is the classification as EPE

$EPE_{s_i,t}$ ($EPE_{s,y,m,d}$) evaluates to 0, if the the true classification of $PE_{s_i,t}$ is the classification as non-EPE.

2.3 Modelling a binary classifier model: The concept of classification logic

Thusfar the prediction of extreme precipitation events has been framed as a binary classification problem, a S.o.T. has been established and a new variable has been defined, by the means of which generated classifications can be evaluated for their correctness. This section then will introduce a concept of modelling a binary classification.

The purpose of classifier models is take a specific input and return a classification as output. In the context of classifying precipitation events, the classifier model of interest need to return a binary classification of precipitation events as either extreme or non-extreme. After training the classifier is completed, it will have classification logic of the form

$$\widehat{EPE}_{s_i,t} = \begin{cases} 1, & \text{condition A} \\ 0, & \text{condition B (otherwise)} \end{cases} \quad (2.6)$$

$$(2.7)$$

where

$\widehat{EPE}_{s_i,t}$ is the generated classification of $PE_{s_i,t}$

$\widehat{EPE}_{s_i,t}$ evaluates to 1, if condition A is true

$\widehat{EPE}_{s_i,t}$ evaluates to 0, if condition B is true

a value of 1 indicates the classification as extreme

a value of 0 indicates the classification as non-extreme

When combined, the conditions 2.6 and 2.7 make the individual classification logic of the respective classifier model. These conditions involve the spatio-temporal location of precipitation events as model input (predictor). Other than that however there is a myriad of possible way to define condition A and condition B. One such way is by the means of a self-contained model. For such self-contained models, one broad distinction can be made between deterministic and statistical models, and between scoring models and probability estimation models. IDW based interpolation in fact is a deterministic scoring method, while a trained GAMM is a statistical probability estimation model.

The following chapter shortly characterizes deterministic scoring models and statistical probability estimation models. In doing so, it concludes the first step of answering the research question, which is the framing of the prediction of extreme precipitation as a classification decision. That chapter is followed by the second step of the main contribution of this thesis, which goes more into detail of how IDW based interpolation and a trained GAMM are appropriate alternatives for defining the classification of precipitation events as extreme and non-extreme. The drawback of such exhaustive focus on the autocorrelation structure is the negligence of other widely accepted explanatory variables as have been described above.

2.4 Defining a classification logic based on deterministic and statistical models

Essentially, regression is a powerful tool, whenever objective is to cover up for our lack of knowledge of the exact dynamics of the data-generating process responsible for the observations we make through the utilization of statistical distributions (cf. Zumel and Mount [2014])(cf.

Wikle et al. [2019], p. 6 and pp. 78)(cf. Fahrmeir et al. [2021], pp. 23). Such is the case with precipitation, where even the most complex models fail at times. It does however inherently follow necessarily, that statistical models are fundamentally superior to deterministic modelling approaches. IDW based interpolation for example is among the simplest deterministic approaches for modelling precipitation, while still accounting for a key feature of its distribution. The choice of this paper, to provide a competitive evaluation of the performance of both IDW based interpolation as a representative of a deterministic modelling approach and training a GAMM as a representative of a statistical (regression) approach, reflect this belief. There are strengths and weakness to both which correspond to their respective fundamental conceptual assumption:

In contrast to a statistical model, a deterministic model does not include any random variation. Neither the response nor the predictor variables are assumed to be randomly distributed and error variance and the uncertainty of model predictions are non-measurable as the direct consequence. A deterministic scoring model then predicts a score, based on a specific dependency of the model response on either other predictors or on itself, which are also scores. To simplify the conceptualization of a scoring model, a score in this context is considered a measurable feature, that is not a probability⁸. A major advantage of deterministic models are, that researchers do not need to inconvenience themselves with statistics (the fact that deterministic models can become quite complex notwithstanding). A researcher may just infer a mechanism from his or her own observations and try to approximate the true process by error- and trial, as is common in the natural sciences. The major drawback is the lack of a way to obtain an intuition for the uncertainty of the predictions. Prediction error may be measured; it is not possible though to estimate such error for future predictions.

(cf. Zumel and Mount [2014])(cf. Wikle et al. [2019], p. 6)(cf. Fahrmeir et al. [2021], pp. 23) The the statistical model does not have these problems. Random variation is present in the form of model response and predictors, such that either some or all are assumed to follow random distributions. For these models, measuring error variance and prediction uncertainty is possible and even an important step towards inference. A statistical probability estimation model predicts a probability of the occurrence of an event, and in the context of the intended use of the model for

⁸In that respect, the thesis differs from Zumel and Mount [2014], where probabilities are included in the definition of scores.

the definition of conditions A(see 2.6) and B(see 2.7), such an event can be seen as the occurrence of the classification of a precipitation event as extreme. The estimated probability is therefore $p(PE_{s_i,t} \text{ is an EPE})$ while $1 - p$ then describes the probability of the classification of a precipitation events as non-extreme. The model response, which is ultimately modelled while training the model to predict probabilities of interests, consists either of the estimations for exactly these probabilities of $PE_{s_i,t}$ being extreme $\forall s, t \in ST$ in the training data, or of a binary variable, that directly indicates the classification of a precipitation events as extreme or non-extreme. These advantages of the probability estimation model come at a price, which is the statistical sophistication required for its application.

On the basis of the idea to model the classification logic on the basis of a self-contained model, the binary classifier can now be expressed as

$$\widehat{\text{EPE}}_{s,t} = \begin{cases} 1, & \text{if } f(x_{s,t}) > \theta \\ 0, & \text{if } f(x_{s,t}) \leq \theta \end{cases} \quad (2.8)$$

where

$\widehat{\text{EPE}}_{s,t}$ is the estimated binary classification decision

a value of 1 indicates, that the estimated classification of $PE_{s_i,t}$ is a classification as EPE

a value of 0 indicates, that the estimated classification of $PE_{s_i,t}$ is a classification as non-EPE.

Compared to 2.6 and 2.7, in 2.8 and 2.9 “condition A” and “condition B” have been expressed explicitly with logical expressions that involve the output of a function $f()$. This function can either be the above mentioned deterministic scoring model, in which case $f(x_{s,t}) = \widehat{PH}_{s,t}$ and where $\widehat{PH}_{s,t}$ is the estimated $PH_{s,t}$. If the function $f()$ is the statistical probability estimation model, then were $f(s, t) = \widehat{p}(PE_{s,t} \text{ is in class EPE})$, while $\widehat{p}(PE_{s,t} \text{ is in class EPE})$ is the estimated probability of a classification as EPE. The logical expression is a comparison of the respective model output with a threshold value θ and θ is a specific precipitation height in the case of the

scoring model and it is a probability in case of the probability estimation model. The classifier model can be described thus as a two layered machine: A core model $f()$ projects an observation associated with the spatio-temporal location s, t onto either $\widehat{PH}_{s,t}$ or onto $\widehat{p}(PE_{s,t} \text{ is in class EPE})$, which are then compared with the respective threshold value. The classification finally depends on which of the two logical expressions in 2.8 and 2.9 evaluates to *true* for the given $PE_{s,t}$.

3 Appropriate choices for the classification logic

The following section presents IDW based interpolation and a trained GAMM as concrete approaches to such definition of the classification logic. The general argument is, that IDW based interpolation is a deterministic scoring model which can be utilized in the definition of the classification logic appropriately. It does so, as it accounts for an essential feature of precipitation, i.e. the spatio-temporal autocorrelation of precipitation height. It is also argued, that a GAMM is a statistical probability estimation model, that not only accounts for the spatio-temporal autocorrelation as well, but also allows the implementation of wideheld beliefs of factors that explain precipitation. These explanations evolve around the water cycle and consider *altitude*, *mean_temperature* and *the spatio-temporal location* as the driving factors of precipitation.

3.1 Principles that govern patterns of precipitation

The spatio-temporal precipitation height is an empirical feature of the metereological phenomenon of precipitation and the scientific community usually discusses the following principles as governing the data-generating process⁹:

The Wather Cyle	Atmospheric Conditions
Topography	Geographic Location

⁹It should be noted, that the idea, that there is a spatio-temporal point-process responsible for virtually any observable data, including precipitation data, is extensively covered by Diggle (see @Diggle.2023). The statistics to act upon that notion however exceeds the current level of statistical education of the author of the thesis.

The water cycle is currently the most fundamental frame for explaining precipitation. It describes the movement of water between the earth's surface and the atmosphere in the form of a cycle, where water first evaporates, condenses into clouds, and when droplets within the clouds finally reach a weight sufficiently great, the droplets fall back onto the earth's surface to return the water. Atmospheric conditions, topography and the geographic location are then tied into this cyclic pattern to explain the individual mechanisms:

While the evaporated water condenses inside higher layers of the sky, air temperature is asserted to positively correlate with the amount of water, that the air can hold. The higher the temperature, the greater the precipitation height. Air temperature is not only a volatile atmospheric condition; researchers have established, that the geographical variation between locations that are closer to and further away from the equator matters systematically with reference to air temperature. As the intensity of the sunlight increases towards the equator, air temperature increases as well.

Topography is in general concerned with the shape of the earth's surface. Meteorologists widely assume, that mountains intervene in the movement of air. The idea is, that wind causes air to rise alongside the mountains and hence to reach higher layers of the atmosphere and that this results in increased precipitation at the windward side of the mountain and relatively less precipitation in on the leeward side of the mountains. This outline of the principles that govern the phenomenon of precipitation is necessarily rudimentary. It shall serve as a gross theoretical background for the following explanation of the choice of concrete predictors of precipitation height.

3.2 Spatio-Temporal Autocorrelation of precipitation heights

Another key characteristic alongside these widespread asserted patterns of precipitation is the strong autocorrelation of the precipitation heights. Such autocorrelation refers to nothing else than the striking similarity of observations of precipitation heights at closeby locations. A different way to phrase this is, that precipitation heights cluster spatio-temporally. In real life, these clusters express as something that can be observed anywhere in the world: On a rainy

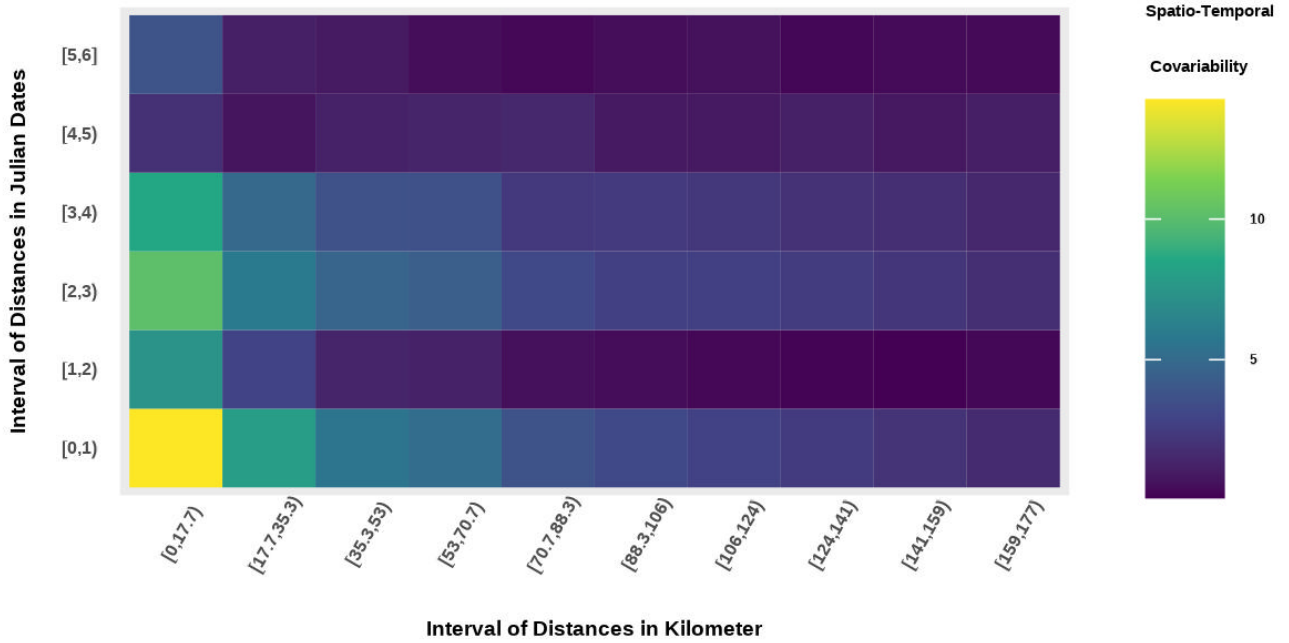


Figure 3: Variation in the Spatio-Temporal Covariance.

day, it will rain or snow in similar amounts up to several kilometers from any spatial locations. And the same logic applies temporally as well. If it rains for instance now, it is more likely that it rained some minutes or hours ago than yesterday, or in some minutes or hours than tomorrow. One can even argue that a human could see a specific manifestation of these clusters in reality, if he or she were to hover in exactly the right distance above the clouds to see where precipitation happens and for how long it rained/snowed. This observation of the spatio-temporal autocorrelation has a profound implication for any attempt to predict extreme precipitation: If a model, that is not capable of accounting for this clustered nature of precipitation, is trained to predict extreme precipitation, it will inevitably generate biased results. They will be biased in the sense, that the prediction error for precipitation in such clusters will be structurally too big, whether the precipitation was over- or underpredicted. While the next sections go on to show how IDW interpolation and GAMMs are capable of handling and identifying such autocorrelation, the existence of such autocorrelation for the precipitation training data will be demonstrated as a preliminary.

The here chosen way to demonstrate the spatio-temporal autocorrelation of precipitation heights in the training data is an analysis of the variation of spatio-temporal covariances. The

corresponding mathematical equation is provided in the appendix (see equation 5.16). The variation is visualized¹⁰ in the figure 3.

(cf. Auer and Rottmann [2020], pp. 90)(cf. Fahrmeir et al. [2021], p.692) Covariances general indicate, whether and how two variables co-behave. Behaviour in this context is to be understood in relation to the variables mean value. For each observation, the difference to the mean of its respective variable is calculated. For a pair of observations of the two variables, the cross-product of these differences are calculated next. If both observations exceed their respective mean, the cross product is positive. And the positive cross-product grows, as the difference the difference of both observations to their respective means grow. Such a cross-product is negative, if one of both observations exceed, while the other deceed their respective mean. The covariance finally is the sum of those cross-products, where a positive value hence indicates, that pairs of observations of both variables tended to mutually exceed or deceed their respective mean, or behave similiarly in other words. And where a negative value indicates, that pairs of observations behaved in exactly the opposite way.

As a concept, spatio-temporal covariance is an extension of this broader concept of measuring similarity in the behaviour of two variables. Instead of assessing how two variables co-behave, it measures, how observations of the same variable co-behave, if they were made at different spatio-temporal locations. It is in this context paramount, that a variety of such covariances are calculated for different groups of observed precipitation heights. These groups can be based on intervals of spatial and temporal distance. All pairs of spatio-temporal locations then for whom the spatial and temporal distances fall inside a pair of spatial and temporal interval would be grouped. Such grouping is necessary, as a single covariance calculated upon the entirety of all pairs of observations couldn't be reasonably interpreted in terms of spatio-temporal autocorrelation: If the covariance for instance is highly positive, it could be, because observations at either close or distant spatio-temporal locations co-behave. If it were highly negative, it could be again, because these paired observations behave opposingly either for close or distant spatio-temporal locations. There would be no way to tell, whether pairs at close or distant locations cause the respective covariance. The only way to utilize covariance appropriately therefore is to group the

¹⁰Visualization was created by the author after calculating the corresponding data in R by means of tidyverse data-manipulation functions. The visualization itself was created by the author applying function of the R package `ggplot2`. The associated R script is available in the GitHub repository.

spatio-temporal observations first. Such grouping was conducted on the base of intervals of spatial and temporal distances. This means: Spatio-temporal observations of precipitation height are paired based on all combinations of spatio-temporal locations. Based on the calculated spatial¹¹ and temporal¹² distances between the paired observations, they are then assigned to intervals of spatial distance and temporal distances. After determining all combinations of intervals, the covariance is calculated only for those pairs of observations within the same combination of intervals of spatial and temporal distance. Any such group of pairs for whom the spatio-temporal distance is within the same duo of distance intervals essentially constitutes a spatio-temporal cluster. An analysis of the variation of the covariance of these clusters finally enable an assessment of the existence of spatio-temporal autocorrelation. If spatio-temporal autocorrelation exists for the distribution of precipitation heights in the training data, then the covariance is expected to decrease

for the same interval of temporal distance with increasing intervals of spatial distance

for the same interval of spatial distance with increasing intervals of temporal distance

along a simultaneous increase of both intervals of spatial and temporal distance.¹³

Considering the visualization of the variation of covariance between the respective interval-based clusters³, all expectations for spatio-temporal autocorrelation are met. The absolute value of the covariance is the greatest for observations that are no more than 17.7 kilometers and no days apart. Aligning with the above defined expectations, the covariance indeed decreases, if either the spatial or temporal interval increases while the other remains the same. The empirical covariance also decreases, if both intervals increase simultaneously.

¹¹The distance between two spatial locations can be measured in different ways. A common way is to calculate the euclidean distance. A drawback of the euclidean distance is, that it assumes a flat space. The earth's surface is characterized by curvature however. Hence a better measure was identified in the Haversine Formula. Using the Haversine Formula, the spatial distance between two spatial locations was measured in kilometer while accounting for the curvature of the earth's surface.

¹²The temporal distance between two observations is measured as the difference between their respective julian dates.

¹³It should be noted, that it does not matter whether the covariance is negative or positive. Both reflect a pattern of behaviour and the higher the absolute value, the greater the similarity of behaviour or cohesiveness of the cluster in other words. What really matters is the difference between clusters of close locations and clusters of distant locations. The absolute values have thus been visualized instead of the original positive or negative values.

The existence of spatio-temporal autocorrelation is therefore safe to assume and the following sections show how it can be accounted for when defining the classification logic.

3.3 IDW based interpolation as an appropriate choice for the classification logic

Everything is related to everything else, but near things are more related than distant things.

- Waldo Tobler [1970]

In contrast to regression, Inverse Distance Weighting based interpolation 5.12 does not model the conditional expected value of a variable like precipitation height, nor the conditional probability for a specific observation to be a member of a certain class. (cf. Wike et al. [2019] , pp. 78)(cf. Bivand et al. [2013] , pp. 215) It models the variable based on a characteristic of its own univariate distribution: its neighborhood in terms of spatial and temporal distance. For any spatio-temporal location, the existing available observations are evaluated against their relative spatial and temporal distance to the spatio-temporal location, for whom the precipitation height is to be interpolated. The model prediction of the variable for that respective location is then determined as a weighted average of all available observed values minus any existing observations at the location in question.

Inverse Distance Weighting based interpolation is a representative of a deterministic scoring model. It assumes, that an observation of a score like precipitation height at any spatio-temporal location can be calculated as an average of all other observations, but such, that observations that are relatively distant have less influence on the predicted value than close observations. The quote cited above is a statement of Waldo Tobler, widely known as *Tobler's first law of geography* and Inverse Distance Weighting based interpolation fundamentally can be considered a practical implication of that notion. As Tobler's law then is one way to approximate the spatio-temporal autocorrelation of precipitation height, which was extensively demonstrated above. Hence, within limitation, it can be appropriately involved in the definition of the classification logic¹⁴.

¹⁴Future research may test this assumption more in-depth. In the context of this thesis, this reasoning of the appropriateness of IDW based interpolation as a classifier core shall suffice.

3.4 A trained GAMM as an appropriate choice for the classification logic

The Generalized Additive Mixed Model with a Logit Link function, as expressed in the subsection above, relies on an algorithm to identify the dependence of precipitation height on mean temperature, altitude, spatial and temporal location. This dependence is broadly speaking expressed as the estimated coefficients and respective smoothing terms. As described, it is indeed suitable for picking up on spatial and temporal dependencies due to its unique modelling concept. However, the mathematical foundation of the formulation and estimation of a GAMM is not as easily accessible as another non-statistical, deterministic interpolation method called “Inverse Distance Weighting, which in comparison requires substantially less statistical expertise.

3.4.1 Classical Linear Regression

Classical linear regression involves modeling the model response as the conditional expected value of a variable, with this expected value being dependent on one or more covariates in a linear fashion, plus an error term.¹⁵ The random error terms are assumed to be identically and independently Gaussian-distributed. The covariates form a linear predictor such that, based on the modeled dependence of the response on both the random errors and the linear predictor, the response is conceptualized as a random variable as well, whose distribution, beyond adopting the i.i.d. features, is additionally influenced by the linear predictor. Importantly, as the error terms conceptually account for random, unsystematic differences between the response and the linear predictor, as well as measurement errors, they are assumed to be independent not only of themselves but also of the linear predictor. These characteristics collectively render linear regression fundamentally inadequate for the present purpose. Firstly, the response is assumed to be Gaussian-distributed, whereas for a binary response, such as a classification decision derived from a probability threshold layer, a binomial distribution needs to be assumed. Secondly, strong evidence exists for the existence of spatial, temporal, and spatio-temporal dependencies between precipitation heights and within the conditional dependency of the precipitation height of the covariates on one side

¹⁵For the model equation, see equation 5.1

and the assumed independence among the responses and among the error terms on the other side. If the responses (the classification decisions) are non-independent (autocorrelation), then these dependencies need to be captured through the linear predictor or the error terms. As the linear predictor, however, is inadequate for modeling any other than linear effects of the covariates on the responses, these dependencies would conceptually be assumed to be reflected in the error terms. The dependencies would therefore reflect in the error terms, particularly as autocorrelation of the error terms, which contradicts the assumption of i.i.d. error terms. Hence, as linear regression is, by construction, inadequate to meet the requirements of modeling a binary response and modeling spatio-temporal dependencies, and as this inadequacy finally also extends to modeling non-linear dependencies, linear regression overall is ill-suited for modeling the classification of PE as EPE, as this thesis is set out for.

3.4.2 Logistic Regression Models as *Generalized Linear Models*

The classical linear regression model can still generate predictions from the spatial and temporal location of a PE. And with changes to the model structure, it can be transformed into a probability estimation model. Although this logistic regression model does not model the binary variable of classification decisions directly, it generates estimations for the probabilities, that a PE is classified as an EPE. To take the final step to model the binary classification variables finally as a model response that depends on a linear predictor then would only require superimposing a probability-threshold layer onto the predictions of the logistic regression model.

But to achieve such a generalization of the linear model towards the logistic regression model, firstly the error term is dropped from the right-hand side, so that compared to linear regression, the model response is no longer modeled as dependent on the linear predictor plus an error term, but only on the linear predictor.

The core acts of the generalization then, which effectively allow to model classification probabilities $p(PE_{s,t} \text{ in class } EPE) = p_{s,t}$ as dependent on a linear predictor, involve

- observations of these probabilities in the data
- a logit-function $\text{logit}(p_{s,t}) = \log\left(\frac{p_{s,t}}{1-p_{s,t}}\right)$, which projects the observed probabilities onto log-odds(logit) and

- a sigmoid function S of the form $S(\text{logit}(p_{s,t})) = \frac{1}{1+e^{\text{logit}(p_{s,t})}} = p_{s,t}$, that is the inverse of the logit-function and returns probabilities with log-odds as input

and the concept is as follows:

It need to be noted first, that the primary concern with modeling a binary response per linear regression had been, that it results in systemic error. The logit-function solves this particular problem by projecting the probability values on a continuous scale from $-\infty$ to ∞ , such that the resulting continuous log-odds can be modeled as dependent on a linear predictor. Since the log-odds are not the variable of interest however, the logit's inverse function, S , is applied on both sides of the model equation. As a consequence, within the logistic regression model, the variable of interest, the probabilities $p_{s,t}$ are modeled as dependent on the inverse logit function S with the logg-odds as the model argument. And since the log-odds were modeled as dependent on a linear predictor in the first place, the model equation presents itself as

$$\begin{aligned}
 p(PE_{s,t} \text{ in class EPE}) &= S\left(\log\left(\frac{p_{s,t}}{1-p_{s,t}}\right)\right) \\
 &= S\left(\begin{aligned} &\beta_0 + \beta_1 \text{Mean Temperature}_{s,t} + \beta_2 \text{Altitude}_{s,t} \\ &+ \beta_3 \text{Longitude}_{s,t} + \beta_4 \text{Latitude}_{s,t} \\ &+ \beta_5 \text{Julian Date}_{s,t} \end{aligned}\right) \quad (3.1)
 \end{aligned}$$

16

3.4.3 Generalized Additive Mixed Model with a logit-link function

Presently, following the above mentioned steps of generalization and on a conceptual level, the spatio-temporal classification of PE as EPE is arguably already partially modelled: The Logistic Regression Model has a probability value as its response after the superimposition of a threshold-layer, the classification of the PE as EPE can be modeled as depending on a linear predictor consisting out of covariates such as altitude, mean-temperature, spatial location and temporal location, while the linear predictor is supplied to the inverse of the logit link-function. From

¹⁶For a in-depth breakdown of the model equation and the concept of logistic regression, see the Appendix at [5.2](#)

here, the remaining two requirements, the model's ability to capture spatial, temporal and spatio-temporal dependencies as well as capture non-linear relations between response and predictor, are introduced in two simultaneous steps: Another generalization, namely the substitution of the linear effects through smoothing-terms, as well as the introduction of random effects. As a consequence, the originally linear predictor is no longer linear and since the smoothing terms are still additively concatenated, ultimately a Generalized Additive Mixed Model with a logit link-function and smoothing terms for effects is built. As this GAMM is able to model the probability of a PE being an EPE as being spatio-temporally dependent on the mentioned covariates, it takes the superimposition of a filtering probability threshold-layer to produce a classifier model that finally meets all the discussed requirements.

$$\begin{aligned} \text{logit}(p(PE_{s,t} \text{ is in class EPE})) &= \beta_0 + f_{lon}(lon_{s,t}) + f_{lat}(lat_{s,t}) \\ &\quad + f_{jul}(jul_t) + f_{mean_temp}(mean_temp_{s,t}) \\ &\quad + f_{altitude}(altitude_{s,t}) + b_{s,t} \end{aligned} \quad (3.2)$$

where $\text{logit}(p_{s,t}) := \ln\left(\frac{p_{s,t}}{1-p_{s,t}}\right)$ is the logit function of the predicted probability of an event $PE_{s,t}$ belonging to the class EPE, $b_{s,t}$ is a random effect term to account for unobserved heterogeneity in the data, and $f_{lon}, f_{lat}, f_{jul}, f_{mean_temp}$, and $f_{altitude}$ are smooth functions of the corresponding predictor variables.

$$p_{s,t} = \frac{1}{\left(1 + \exp(-(\beta_0 + f_{lon_{s,t}}(lon_{s,t}) + f_{lat}(lat_{s,t}) + f_{jul}(t) + f_{mean_temp}(mean_temp_{s,t}) + f_{altitude}(altitude_{s,t})))\right)} \quad (3.3)$$

where $lon_{s,t}, lat_{s,t}, jul_{s,t}, mean_temp_{s,t}$, and $altitude_{s,t}$ are the values of the longitudinal degree, latitudinal degree, julian date, mean daily temperature, and altitude predictor variables for precipitation event $PE_{s,t}$, and f_{lon} to $f_{altitude}$ are smooth functions of these variables.

To convert this GAMM into a classifier model, we add a threshold layer, which classifies events as belonging to the class if their predicted probability of belonging to the class exceeds the threshold:

The classifier model with the threshold layer can be expressed as:

$$\widehat{EPE}_{s,t} = \begin{cases} 1, & \text{if } p_{s,t} > \theta \\ 0, & \text{otherwise} \end{cases} \quad (3.4)$$

where \widehat{EPE} is the predicted class label for event $PE_{s,t}$, and θ is the chosen threshold value.

4 A competitive evaluation of the performance of the classifier models

Up to this point, the prediction of extreme precipitation was framed as the classification of precipitation events as either extreme or non-extreme. It has been established as well, that patterns of precipitation are governed by spatio-temporal autocorrelation and the dependency of precipitation height on altitude, mean temperature and the spatio-temporal location. And IDW based interpolation and a trained GAMM have finally been asserted as appropriate choices when attempting to define the logic behind the classification of precipitation events, precisely because they account for these assumed patterns.

All of these considerations eventually prepare what is to come now, that is, the essence of the answer to the research question, *how the classifier models compare performance-wise..* At the center of this answer are measurements of confusion-matrix based performance indicators. The following three subsections make therefore the last step of the main part of this thesis and they are structured as follows:

Firstly, the concept of confusion-matrix based performance indicators are explained in the context of the above mentioned Standard of Truth. Then, the measured values for the performance-indicators are presented for both classifier models individually. Finally, a comparison of the measurements is conducted and interpreted.

4.1 How confusion-matrix based performance indicators quantify classifier performance

In the a previous section, the relevance of the S.o.T. for evaluative purposes had been established. In this context, the determination of a set of true classifications had been proposed and described (cf. equation 2.5). These true classifications were finally asserted a necessary requirement in practically assessing the truth of a generated classification by the means of a comparison of generated classification and true classification.

(Hosmer et al. [2013], pp. 169)(Zumel and Mount [2014]) For the evaluation of classifier model a general rule applies in fact, whereafter a classifier is better, the more true classifications it generates. A more differentiated outline of the classifier model's performance can be obtained, if the number of true and false classifications are counted while distinguishing between classifications of the $PE_{s_i,t}$ as extreme or non-extreme for $\forall (s_i, t) \in ST$. To provide enhanced readability, and also to align the notation with general practices in the (data) scientific literature, a classification of $PE_{s_i,t}$ as extreme is denoted a *positive*, while a classification of $PE_{s_i,t}$ as non-extreme is denoted a *negative*. By counting the number of true negatives (TN), false positive (FP), false negatives (FN) and true positives (TP) and aligning them in the cells of a 2×2 matrix, the socalled standard 2×2 confusion matrix is obtained, such that

$$\text{Confusion Matrix} := \begin{array}{c|cc} & \text{Predicted Positive} & \text{Predicted Negative} \\ \hline \text{Negative by S.o.T.} & \text{True Positives} & \text{False Negatives} \\ \hline \text{Positive by S.o.T.} & \text{False Positives} & \text{True Negatives} \end{array} \quad (4.1)$$

Now, let Q be the complete set of true classifications of $PE_{s_i,t} \forall (s_i, t) \in ST$ and let P be the complete set of generated classifications of $PE_{s_i,t} \forall (s_i, t) \in ST$. Also let q be a subset of Q and p be a subset of P .

Then:

$$\begin{aligned}
\text{True Positives} &:= &^{17} & \text{False Positives} := \\
|\{p \times q \mid q \in Q, p \in P, q = p = \text{positive}\}| & & |\{p \times q \mid q \in Q, p \in P, q = \text{negative}, p = \text{positive}\}| \\
& & (4.2)
\end{aligned}$$

$$\begin{aligned}
\text{False Negatives} &:= & \text{True Negatives} &:= \\
|\{p \times q \mid q \in Q, p \in P, q = \text{positive}, p = \text{negative}\}| & & |\{p \times q \mid a \in A, p \in P, q = p = \text{negative}\}| \\
& & (4.3)
\end{aligned}$$

As mentioned earlier, these count values or values in the cells of the confusion matrix respectively, are used to acquire a more differentiated view of the performance of the classifier model. Specifically they are the basis for the calculation of a set of performance indicators, called *Precision*, *Sensitivity*, *Specificity*, *False Positive Rate* and *False Negative Rate*.¹⁸ Their respective formulae are as follows:

$$\begin{aligned}
\text{Precision} &:= \frac{TP}{TP + FP} & \text{Sensitivity} &:= \frac{TP}{TP + FN} & \text{Specificity} &:= \frac{TN}{TN + FP} \\
\text{False Positive Rate} &:= \frac{FP}{TN + FP} & \text{False Negative Rate} &:= \frac{FN}{TP + FN}
\end{aligned} \tag{4.4}$$

Each of these measure a particular aspect of the classifier performance:

The **Precision** of a classifier is the fraction of the generated positives, for which the generated classification as positive is correct. This fraction gives an answer to the question *how often a generated classification of $PE_{s_i,t}$ as extreme is true*.

The **Sensitivity** of a classifier is the fraction of the true generated classifications of $PE_{s_i,t}$ as extreme to the overall number of true positives. It gives an answer to *how many times the classifier correctly identified extreme precipitation events*.

¹⁷On the notation: $|p \times q|$ refers to the number of elements of " $p \times q$ ", while $p \times q$ is the cartesian product of the two subsets p and q

¹⁸Another indicator called *Accuracy* is not considered. The problem with Accuracy, which describes the ratio of predicted true classifications to the number of predicted classifications, is a strong bias in the case of asymmetrical distributions, as is the case with Precipitation events, where most of the precipitation events are non-extreme.

The **Specificity** of a classifier is the fraction of the true classifications of precipitation events as non-extreme to the number of true classifications of PE as non-extreme in the training data. It gives an answer to *how many times the classifier correctly identified non-extreme precipitation (or precipitation events as non-extreme)*.

The **False Positive Rate** is the performance indicator complementary to the True Negative Rate, as it describes the fraction of true classifications of precipitation events as non-extreme in the training data, which the model didn't generate.

The **False Negative Rate** is the indicator of model performance, which is complementary to the true Positive Rate, as it describes the fraction of true classifications of precipitation events as extreme in the training data, which the model didn't generate.

4.2 Confusion-matrix and performance indicators for the IDW based classifier

4.3 Consufion-matrix and performance indicators for the GAMM based classifier

4.4 A comparison of the performance indicators of both classifiers

5 Conclusion

Appendix

The appendix serves as a succinct reference to the mathematical structure of both models, that were described in the context of defining the classification logic and also of measurements of autocorrelation.

Defining the classification logic

In this subsection of the appendix, models are presented mathematically that were previously referenced in the context of defining the classification logic.

Classical Linear Regression

$$\begin{aligned} \text{Precipitation Height} = & \beta_0 + \beta_1 \text{Mean Temperature} + \beta_2 \text{Altitude} \\ & + \beta_3 \text{Longitude} + \beta_4 \text{Latitude} + \beta_5 \text{Julian Date} + \epsilon \end{aligned} \tag{5.1}$$

with assumptions

Linearity:

$$y_{s,t} = \beta_0 + \sum_{j=1}^p \beta_j x_{j,s,t} + \epsilon_{s,t}$$

Independence:

$$E(\epsilon_{s,t}) = 0, \quad Var(\epsilon_{s,t}) = \sigma^2, \quad Cov(\epsilon_{s_i,t_i}, \epsilon_{(s_j,t_j)}) = 0 \quad \forall (s_i,t_i) \neq (s_j,t_j)$$

Homoscedasticity:

$$Var(\epsilon_{s,t}) = \sigma^2 \quad \forall s, t \in ST$$

Normality:

$$\epsilon_{s,t} \sim N(0, \sigma^2)$$

No perfect multicollinearity:

$$\text{rank}(X) = p$$

No autocorrelation:

$$\text{Cov}(\epsilon_{s_i, t_i}, \epsilon_{s_j, t_j}) = 0 \quad \forall \quad \begin{pmatrix} s_i \\ t_i \end{pmatrix} \neq \begin{pmatrix} s_j \\ t_j \end{pmatrix}$$

Declaration of Symbols and Equation Components

Symbols related to equation [5.1](#)

Precipitation Height: Response

β_0 : Intercept parameter.

β_1 : Coefficient for *MeanTemperature*.

β_2 : Coefficient for *Altitude*.

β_3 : Coefficient for *Longitude*.

β_4 : Coefficient for *Latitude*.

β_5 : Coefficient for *JulianDate*.

ϵ : Error term.

Symbols related to the assumptions

Linearity Assumption:

$y_{s,t}$: Response.

β_j : Coefficient for the j -th predictor variable.

$x_{j,s,t}$: Value of j -th predictor variable at space-time location s, t .

Independence Assumption:

$E(\epsilon_{s,t})$: Expected value of error term for precipitation height observed at space-time location.

$Var(\epsilon_{s,t})$: Variance of error term at space-time location.

$Cov(\epsilon_{s_i,t_i}, \epsilon_{s_j,t_j})$: Covariance of error terms between two space-time locations.

Homoscedasticity Assumption:

σ^2 : Constant variance of error term.

$s, t \in ST$: Space-time locations as elements of the space-time domain derived from the combination of the two-dimensional spatial domain (longitude,latitude) and the temporal domain made up of julian dates.

Assumption of normality:

$\epsilon_{s,t}$: Error term.

$N(0, \sigma^2)$: Normal distribution with mean 0 and variance σ^2 .

Assumption of no perfect multicollinearity :

$\text{rank}(X)$: Rank of the design matrix X .

p : Number of predictor variables.

Assumption of absend autocorrelation:

$Cov(\epsilon_{s_i,t_i}, \epsilon_{s_j,t_j})$: Covariance of error terms.

5.0.1 Logistic Regression

$$\begin{aligned}
 p(PE_{s,t} \text{ in class EPE}) &= \text{S} \left(\log \left(\frac{p_{s,t}}{1 - p_{s,t}} \right) \right) \\
 &= \text{S} \left(\begin{aligned} &\beta_0 + \beta_1 \text{Mean Temperature}_{s,t} + \beta_2 \text{Altitude}_{s,t} \\ &+ \beta_3 \text{Longitude}_{s,t} + \beta_4 \text{Latitude}_{s,t} \\ &+ \beta_5 \text{Julian Date}_{s,t} \end{aligned} \right) \quad (5.2)
 \end{aligned}$$

where

Coefficients

$$\beta_0, \beta_1, \beta_2, \beta_3, \text{ and } \beta_4 \quad (5.3)$$

are *estimated* per *Maximum Likelihood Estimation*.

Predicted Probability

$$\begin{aligned}
 p_{s,t} &= \left(1 + e^{-(\beta_0 + \beta_1 \text{Mean Temperature}_{s,t} + \beta_2 \text{Altitude}_{s,t} + \beta_3 \text{Longitude}_{s,t})} \right. \\
 &\quad \left. \times e^{-(\beta_4 \text{Latitude}_{s,t} + \beta_5 \text{Julian Date}_{s,t})} \right)^{-1} \quad (5.4)
 \end{aligned}$$

is the model prediction of the probability that the PE at spatio-temporal location s, t is an

EPE.

Conditional Probability

$$P(EPE_{s,t}|\mathbf{x}_{s,t}, \beta) = p_{s,t}^{EPE_{s,t}}(1 - p_{s,t})^{1-EPE_{s,t}} \quad (5.5)$$

is the conditional probability of observing an EPE at the given location and time, given the predicted probability $p_{s,t}$. The term $p_{s,t}^{EPE_{s,t}}$ represents the probability of observing that an extreme precipitation event actually occurs at the spatio-temporal location, and $(1 - p_{s,t})^{1-EPE_{s,t}}$ represents the probability of not observing an EPE at the spatio-temporal location.

Likelihood

$$L(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5) = \prod_{s,t \in SP} P(EPE_{s,t}|\mathbf{x}_{s,t}, \beta) \quad (5.6)$$

$$= \prod_{s,t \in SP} p_{s,t}^{EPE_{s,t}}(1 - p_{s,t})^{1-EPE_{s,t}} \quad (5.7)$$

is the likelihood function or the likelihood of the entire training data, respectively.

Coefficient Estimate per Iteratively Reweighted Least Squares

$$\beta_c = \min_{\beta} \sum_{s,t \in SP} w_{s,t,c} \left(EPE_{s,t} - x_{s,t}^T \beta \right)^2 \quad (5.8)$$

is the estimate of the β through *Iteratively Reweighted Least Squares* (cf. Wood [2017] pp.249)(cf. Fahrmeir et al. [2021] pp.193)

5.0.2 Generalized Additive Mixed Model

$$\begin{aligned}
\text{logit}(p(PE_{s,t} \text{ is in class EPE})) &= \beta_0 + f_{lon}(lon_{s,t}) + f_{lat}(lat_{s,t}) \\
&+ f_{jul}(jul_t) + f_{mean_temp}(mean_temp_{s,t}) \\
&+ f_{altitude}(altitude_{s,t}) + b_{s,t}
\end{aligned} \tag{5.9}$$

where $\text{logit}(p_{s,t}) := \ln\left(\frac{p_{s,t}}{1-p_{s,t}}\right)$ is the logit function of the predicted probability of an event $PE_{s,t}$ belonging to the class EPE, $b_{s,t}$ is a random effect term to account for unobserved heterogeneity in the data, and $f_{lon}, f_{lat}, f_{jul}, f_{mean_temp}$, and $f_{altitude}$ are smooth functions of the corresponding predictor variables.

$$p_{s,t} = \frac{1}{\left(1 + \exp(-(\beta_0 + f_{lon_{s,t}}(lon_{s,t}) + f_{lat}(lat_{s,t}) + f_{jul}(t) + f_{mean_temp}(mean_temp_{s,t}) + f_{altitude}(altitude_{s,t})))\right)} \tag{5.10}$$

where $lon_{s,t}$, $lat_{s,t}$, $jul_{s,t}$, $mean_temp_{s,t}$, and $altitude_{s,t}$ are the values of the longitudinal degree, latitudinal degree, julian date, mean daily temperature, and altitude predictor variables for precipitation event $PE_{s,t}$, and f_{lon} to $f_{altitude}$ are smooth functions of these variables.

To convert this GAMM into a classifier model, we add a threshold layer, which classifies events as belonging to the class if their predicted probability of belonging to the class exceeds the threshold:

The classifier model with the threshold layer can be expressed as:

$$\widehat{EPE}_{s,t} = \begin{cases} 1, & \text{if } p_{s,t} > \theta \\ 0, & \text{otherwise} \end{cases} \tag{5.11}$$

where \widehat{EPE} is the predicted class label for event $PE_{s,t}$, and θ is the chosen threshold value.

Inverse Distance Weighting Based Interpolation

$$\widehat{rain}_{s_i, t_i} := \sum_{s_j, t_j \in ST} \left(w_{(s_i, t_i), (s_j, t_j)} \cdot rain_{s_j, t_j} \right) \quad (5.12)$$

(cf. Wikle et al. [2019] , pp. 78)(cf. Bivand et al. [2013] , pp. 215)

where

$$d_{(s_i, t_i), (s_j, t_j)}^\alpha := \left(\sqrt{(lon_{s_i} - lon_{s_j})^2 + (lat_{s_i} - lat_{s_j})^2 + C(t_i - t_j)^2} \right)^\alpha \quad (5.13)$$

and

$$w_{(s_i, t_i), (s_j, t_j)} := \frac{d_{(s_i, t_i), (s_j, t_j)}^{-\alpha}}{\sum_{s_j, t_j \in ST} d_{(s_i, t_i), (s_j, t_j)}^{-\alpha}} \quad (5.14)$$

Declaration of Symbols and Equation Components

$\widehat{rain}_{s, t}:$

The estimated precipitation height at a specific location and time, denoted by longitude lon_i , latitude lat_r , and Julian day t .

$\sum_{s_j, t_j \in ST} :$

A summation over all the locations and times in the spatio-temporal domain $\mathbb{R}^2 \times \mathbb{R}$ emerging from the conjunction of the two-dimensional spatial domain, within which longitudinal and latitudinal degree each establish one dimension, and the one-dimensional spatial domain, consisting of julian dates as real numbers.

$$\left(\mathbf{w}_{(s_i, t_i), (s_j, t_j)} \cdot \mathbf{rain}_{(s_j, t_j)} \right):$$

A product of two terms. The first term, $w_{((s_j, t), s, t)}$, is a weight assigned to each location and time in ST based on its distance from s_i, t_i , the location and time of interest. The second term, $rain_{s_j, t_j}$, is the available observed precipitation height at the spatio-temporal location s_j, t_j .

Measurements of spatial, temporal and spatio-temporal autocorrelation

This subsection of the appendix mathematically presents measurements of autocorrelation for reference.

Haversine Formula

$$d := 2r \arcsin \left(\sqrt{\sin^2 \left(\frac{\phi_2 - \phi_1}{2} \right) + \cos(\phi_1) \cos(\phi_2) \sin^2 \left(\frac{\lambda_2 - \lambda_1}{2} \right)} \right) \quad (5.15)$$

Dauni et al. [2019]

Declaration of Symbols and Equation Components

d:

Distance between two spatial points in kilometer

r:

Radius of the world in kilometer (6371km)

ϕ_1 and ϕ_2 :

Latitudinal Degrees of the two points measured in radians. ϕ is calculated from a decima latitudinal degree per $\phi = lat \frac{\pi}{180}$, where lat is a latitudinal degree.

λ_1 and λ_2 :

Longitudinal Degrees of the two points measured in radians. λ is calculated from a decimal latitudinal degree per $\lambda = lon \frac{\pi}{180}$

sin and **cos**:

Trigonometric functions. For a given angle measured in radians, both functions return either the sine or the cosine.

arcsin:

$\arcsin(x)$ is the inverse of the sine function. For a given sine x , it returns the sine, whose angle measured (measured in radians) is then x

Empirical Spatio-Temporal Covariance

$$\hat{C}_{rain}(d_{spatial}, d_{julian}) := \frac{1}{|N(d_{spatial})|} \frac{1}{|N(d_{julian})|} \times \left(\sum_{s_i, s_j \in N(d_{spatial})} \sum_{t_i, t_j \in N(d_{julian})} (rain_{s_i, t_i} - \hat{\mu}_{t_i}) (rain_{s_j, t_j} - \hat{\mu}_{t_j}) \right) \quad (5.16)$$

Cressie and Wikle [2011] | Wikle et al. [2019] | Auer and Rottmann [2020]

Declaration of Symbols and Equation Components

$\hat{C}_{rain}(\mathbf{d}_{spatial}, \mathbf{d}_{julian})$:

The empirical covariance of the precipitation height data for a given spatial distance in kilometer $d_{spatial}$ and a given temporal distance d_{julian} .

$\mathbf{d}_{spatial}$:

The interval of spatial distances between two weather stations denoted each as a pair of angular coordinates, within which the precipitation heights are considered to be correlated.

\mathbf{d}_{julian} :

The interval of temporal distances between t and t , within which the precipitation heights are considered to be correlated

\mathbf{t}_i and \mathbf{t}_j :

Julian days or the days elapsed since the earliest observation within the precipitation data, respectively.

$\mathbf{N}(\mathbf{d}_{\text{spatial}})$:

A set of pairs of weather stations, for whom the spatial distance within the pair evaluates to a value within the interval d_{spatial} .

$\mathbf{N}(\mathbf{d}_{\text{julian}})$:

A set of pairs of weather stations, for whom the temporal distance within the pair evaluates to a value within the interval d_{julian} .

$|\mathbf{N}(\mathbf{d}_{\text{spatial}})|$:

The number of pairs weather stations within the set d_{spatial} .

$|\mathbf{N}(\mathbf{d}_{\text{julian}})|$:

The number of pairs of temporal locations of the weather stations that are within the set d_{julian} .

$$\sum_{(\mathbf{s}_i, \mathbf{s}_j) \in \mathbf{N}(\mathbf{d}_{\text{spatial}})} \sum_{(\mathbf{t}_i, \mathbf{t}_j) \in \mathbf{N}(\mathbf{d}_{\text{julian}})} :$$

A double summation over pairs of observations of the precipitation height made at a pair of weather stations spatio-temporally located, such that the pair belongs to d_{spatial} spatially and d_{julian} temporally.

$\mathbf{rain}_{\mathbf{s}_i, \mathbf{t}_i}$ and $\mathbf{rain}_{\mathbf{s}_j, \mathbf{t}_j}$:

The precipitation height value at spatio-temporal location s_i, t_i and s_j, t_j

$\hat{\mu}_{\text{rain,temporal}}(\mathbf{t}_i)$ and $\hat{\mu}_{\text{rain,temporal}}(\mathbf{t}_j)$:

The spatial mean of the precipitation height data at time t_i and t_j .

$(\text{rain}_{\mathbf{s}_i, \mathbf{t}_i} - \hat{\mu}_{\text{rain,temporal}}(\mathbf{t}_i))$:

The deviation of the precipitation height data from their temporal mean at spatio-temporal location s_i, t_i

$(\text{rain}_{\mathbf{s}_j, \mathbf{t}_j} - \hat{\mu}_{\text{rain,temporal}}(\mathbf{t}_j))$:

The deviation of the precipitation height data from their temporal mean at spatio-temporal location s_j, t_j .

$(\text{rain}_{\mathbf{s}_i, \mathbf{t}_i} - \hat{\mu}_{\text{rain,temporal}}(\mathbf{t}_i))(\text{rain}_{\mathbf{s}_j, \mathbf{t}_j} - \hat{\mu}_{\text{rain,temporal}}(\mathbf{t}_j))$:

The product of the two deviations

$\frac{1}{2|\mathbf{N}(\mathbf{d}_{\text{spatial}})|} \frac{1}{2|\mathbf{N}(\mathbf{d}_{\text{julian}})|}$:

A normalization factor that scales the covariance to account for the number of pairs of weather stations, for whom the distance in kilometer evaluates to a value within d_{spatial} , while their temporal distance evaluates to value within the interval d_{temporal} being used in the calculation.

Empirical Spatial Covariance

$$\hat{C}_{rain, (d_{julian})}(s_i, s_j) := \frac{1}{T - d_{julian}} \times \left(\sum_{t=1+d_{julian}}^T (rain_{s_i, t} - \hat{\mu}_{rain, temporal}(s_i)) \right. \\ \left. (rain_{s_j, t-d_{julian}} - \hat{\mu}_{rain, temporal}(s_j)) \right) \quad (5.17)$$

Cressie and Wikle [2011] | Wikle et al. [2019] | Auer and Rottmann [2020]

Declaration of Symbols and Equation Components

$\hat{C}_{rain, d_{julian}}$:

The estimated spatial covariance of the precipitation data, such that a given temporal distance d_{julian} applies to all pairs of observations, for whom the cross-product is calculated.

s_i and s_j :

Two spatial locations of weather stations.

T :

The number of julian dates or temporal locations, respectively.

d_{julian} :

A specific temporal distance between two julian dates.

$rain_{s_i, t}$ and $rain_{s_j, t}$:

Observations of precipitation height at the spatio-temporal locations s_i, t and s_j, t .

$\hat{\mu}_{\text{rain,spatial}}(\mathbf{s}_i)$ and $\hat{\mu}_{\text{rain,spatial}}(\mathbf{s}_j)$:

The estimated means of the precipitation data at locations \mathbf{s}_i and \mathbf{s}_j , respectively.

Empirical Temporal Covariance

$$\hat{C}_{rain}^{(d_{spatial})}(t, t) := \frac{1}{|N(d_{spatial})|} \times \left(\sum_{(lon_r, lat), (lon_s, lat_k) \in N(d_{spatial})} (rain_{lon_r, lat, t} - \hat{\mu}_{rain, spatial}(t)) \right. \\ \left. (rain_{lon_s, lat_k, t} - \hat{\mu}_{rain, spatial}(t)) \right) \quad (5.18)$$

Cressie and Wikle [2011] | Wikle et al. [2019] | Auer and Rottmann [2020]

Declaration of Symbols and Equation Components

$\hat{\mathbf{C}}_{rain}^{(d_{spatial})}$:

The estimated covariance of the precipitation data for a given spatial distance $d_{spatial}$.

\mathbf{t} and \mathbf{t} :

Julian days that are a certain temporal distance apart.

$\mathbf{N}(\mathbf{d}_{spatial})$:

A set of pairs of spatial locations, for whom the spatial distance between the paired locations is within a certain interval of spatial distance $d_{spatial}$.

$|\mathbf{N}(\mathbf{d}_{spatial})|$:

The number of pairs spatial locations / weather stations, for whom the spatial distance lies within the interval of spatial distance $d_{spatial}$.

$(\mathbf{lon}_r, \mathbf{lat})$:

A pair of longitude and latitude coordinates.

$\mathbf{rain}_{lon_r, lat, t}$:

The precipitation height value at the location (lon_r, lat) and time t .

$\hat{\mu}_{\text{rain,temporal}}(\mathbf{t})$ and $\hat{\mu}_{\text{rain,temporal}}(\mathbf{t})$:

The estimated means of the precipitation height data for times t and t , respectively.

Empirical Spatial Semivariogram

$$\gamma(d_{spatial}, t) := \frac{1}{2|N(d_{spatial})|} \times \left(\sum_{s_i, s_j \in N(d_{spatial})} (rain_{s_i, t} - rain_{s_j, t})^2 \right) \quad (5.19)$$

Cressie and Wikle [2011] | Wikle et al. [2019]

Declaration of Symbols and Equation Components

$\gamma(\mathbf{d}_{spatial}, \mathbf{t})$:

The semivariogram function for the precipitation height data over a given interval of spatia distance in kilometer $d_{spatial}$ on a given day t .

$\mathbf{d}_{spatial}$:

An interval of spatial distance between weather stations measured in kilometer within which the precipitation height data are considered to be correlated.

$\mathbf{N}(\mathbf{d}_{spatial})$:

The set of pairs of weather stations, for whom the distance between them lies within the interval $d_{spatial}$.

$|\mathbf{N}(\mathbf{d}_{spatial})|$:

The number of pairs of weather stations in the set $N(d_{spatial})$.

\mathbf{s}_i and \mathbf{s}_j :

Two spatial locations of a weather stations

$\mathbf{rain}_{\mathbf{s}_i, t}$ and $\mathbf{rain}_{\mathbf{s}_j, t}$:

Observations of precipitation height at the spatio-temporal locations s_i and s_j for the Julian day t .

Empirical Temporal Semivariogram

$$\gamma(d_{julian}, s) := \frac{1}{2|N(d_{julian})|} \times \left(\sum_{(t_i, t_j) \in N(d_{julian})} (rain_{s, t_i} - rain_{s, t_j})^2 \right) \quad (5.20)$$

Cressie and Wikle [2011] | Wikle et al. [2019] | Auer and Rottmann [2020]

Declaration of Symbols and Equation Components

$\gamma(\mathbf{d}_{julian}, \mathbf{s})$:

The temporal semivariogram function for the precipitation height data at the location s over a given interval of temporal distances d_{julian} .

\mathbf{d}_{julian} :

An interval of temporal distances within which the precipitation data are considered to be correlated.

$\mathbf{N}(\mathbf{d}_{julian})$:

The set of pairs of julian days, for whom their distance evaluates to a value within a certain interval d_{julian} .

$|\mathbf{N}(\mathbf{d}_{julian})|$:

The number of pairs of julian days within the set $N(d_{julian})$.

$(\mathbf{t_i}, \mathbf{t_j})$:

A pair of two julian days

$\mathbf{rain_{s,t_i}}$ and $\mathbf{rain_{s,t_j}}$:

Observations of precipitation height at the location s and the julian days t_i and t_j , respectively.

6 References

- Benjamin Auer and Horst Rottmann. *Statistik und Ökonometrie für Wirtschaftswissenschaftler*. Springer Fachmedien Wiesbaden, Wiesbaden, 2020. ISBN 978-3-658-30136-1. doi: 10.1007/978-3-658-30137-8.
- Roger Bivand, Edzer J. Pebesma, and Virgilio Gómez-Rubio. *Applied spatial data analysis with R*. Use R! Springer, New York and Heidelberg and Dordrecht and London, second edition edition, 2013. ISBN 978-1-4614-7618-4.
- Noel A. C. Cressie and Christopher K. Wikle. *Statistics for spatio-temporal data*. Wiley series in probability and statistics. Wiley, Hoboken, N.J, 2011. ISBN 9781119243069. URL <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=4427975>.
- P. Dauni, M. D. Firdaus, R. Asfariani, M. I. N. Saputra, A. A. Hidayat, and W. B. Zulfikar. Implementation of haversine formula for school location tracking. *Journal of Physics: Conference Series*, 1402(7):077028, 2019. ISSN 1742-6588. doi: 10.1088/1742-6596/1402/7/077028.
- Ludwig Fahrmeir, Thomas Kneib, Stefan Lang, and Brian D. Marx. *Regression: Models, Methods and Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2. 2nd ed. 2021 edition, 2021. ISBN 9783662638828. URL <http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1987674>.
- Luis Gimeno, Rogert Sorí, Marta Vázquez, Milica Stojanovic, Iago Algarra, Jorge Eiras-Barca, Luis Gimeno-Sotelo, and Raquel Nieto. Extreme precipitation events. *WIREs Water*, 9(6), 2022. ISSN 2049-1948. doi: 10.1002/wat2.1611.
- David W. Hosmer, Stanley Lemeshow, and Rodney X. Sturdivant. *Applied logistic regression*, volume 398 of *Wiley series in probability and statistics*. Wiley, Hoboken, third edition edition, 2013. ISBN 978-0-470-58247-3. doi: 10.1002/9781118548387. URL <https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=566576>.
- Sina Keller and Andreas Atzl. Mapping natural hazard impacts on road infrastructure—the extreme precipitation in baden-württemberg, germany, june 2013. *International Journal of Disaster Risk Science*, 5(3):227–241, 2014. ISSN 2095-0055. doi: 10.1007/s13753-014-0026-1.

- Heidi Kreibich, Meike Müller, Kai Schröter, and Annegret H. Thielen. New insights into flood warning reception and emergency response by affected parties. *Natural Hazards and Earth System Sciences*, 17(12):2075–2092, 2017. doi: 10.5194/nhess-17-2075-2017.
- Christian Kuhlicke, Ines Callsen, and Chloe Begg. Reputational risks and participation in flood risk management and the public debate about the 2013 flood in germany. *Environmental Science & Policy*, 55:318–325, 2016. ISSN 14629011. doi: 10.1016/j.envsci.2015.06.011.
- Annegret H. Thielen, Sarah Kienzler, Heidi Kreibich, Christian Kuhlicke, Michael Kunz, Bernhard Mühr, Meike Müller, Antje Otto, Theresia Petrow, Sebastian Pisi, and Kai Schröter. Review of the flood risk management system in germany after the major flood in 2013. *Ecology and Society*, 21(2), 2016. ISSN 1708-3087. doi: 10.5751/ES-08547-210251.
- Annegret H. Thielen, Philip Bubeck, Anna Heidenreich, Jennifer von Keyserlingk, Lisa Dillenardt, and Antje Otto. *Performance of the flood warning system in Germany in July 2021 – insights from affected residents*. 2022. doi: 10.5194/egusphere-2022-244.
- W. R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:234, 1970. ISSN 00130095. doi: 10.2307/143141. URL <http://www.jstor.org/stable/143141>.
- S. Trömel and C.-D. Schönwiese. Probability change of extreme precipitation observed from 1901 to 2000 in germany. *Theoretical and Applied Climatology*, 87(1-4):29–39, 2007. ISSN 0177-798X. doi: 10.1007/s00704-005-0230-4.
- Dennis Wagenaar, Stefan Lüdtke, Kai Schröter, Laurens M. Bouwer, and Heidi Kreibich. Regional and temporal transferability of multivariable flood damage models. *Water Resources Research*, 54(5):3688–3703, 2018. ISSN 0043-1397. doi: 10.1029/2017WR022233.
- Christopher K. Wikle, Andrew Zammit Mangion, and Noel A. C. Cressie. *Spatio-temporal statistics with R*. Chapman & Hall/CRC. CRC Press Taylor & Francis Group, Boca Raton and London and New York, 2019. ISBN 978-1-138-71113-6.
- Simon N. Wood. *Generalized additive models: An introduction with R*. Texts in statistical science. CRC Press Taylor & Francis Group, Boca Raton and London and New York, second edition

edition, 2017. ISBN 9781498728348. URL <https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=4862399>.

Nina Zumel and John Mount. *Practical data science with R*. Manning, Shelter Island, 2014. ISBN 9781617291562.