

Universität Koblenz
Fachbereich 3: Mathematik/Naturwissenschaften
Geographisches Institut
Biologisches Institut
Bundesanstalt für Gewässerkunde

Biogeomorphological correlations considering aggradation and erosion along the River Elbe using the R programming language

Masterarbeit

zur Erlangung des akademischen Grades

„Master of Science“

vorgelegt am 05. August 2021 von:

Christopher Nicholls

Viktoriastr. 23

56068 Koblenz

Matrikel-Nr.: 219201980

cnicholls@uni-koblenz.de

Gutachter:

Dr. Arnd Weber,

Prof. Dr. Klaus Fischer

Dr. Michael Tempel

Statement of Authentication

I *Christopher Nicholls* hereby declare that I have written the present thesis titled

“Biogeomorphological correlations considering aggradation and erosion along the River Elbe using the R programming language”

independently, without assistance from external parties, and without use of other resources than those indicated. The ideas taken directly or indirectly from external sources (including electronic sources) are duly acknowledged in the text. The material, either in full or in part, has not been previously submitted for grading at this or any other academic institution.

Koblenz, 03.08.2021

Place, Date



Author signature

Verfassererklärung

Hiermit erkläre ich *Christopher Nicholls* an Eides statt, dass ich die vorliegende Arbeit mit dem Titel

“Biogeomorphological correlations considering aggradation and erosion along the River Elbe using the R programming language”

selbständig, ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen (einschließlich elektronischer Quellen) direkt oder indirekt übernommenen Gedanken sind ausnahmslos als solche kenntlich gemacht. Die Arbeit ist in gleicher oder ähnlicher Form oder auszugsweise im Rahmen einer anderen Prüfung noch nicht vorgelegt worden.

Koblenz, 03.08.2021

Ort, Datum



Unterschrift der Verfasserin

Abstract

This work employed integrated nested Laplace approximation (INLA) to statistically analyse the most important factors influencing the annual erosion rate of River Elbe in Germany. This was achieved by using the R package INLA developed by Rue et al. (2009). Via INLA and a fork of the function `INLAsstep()` from the R package `INLAutils` (Redding et al. 2019) a total of 67 environmental parameters were fit to the annual erosion rates from the three different time periods 1898–1995, 1995–2004, and 2004–2017. A total of 6017 model formulas were tested for each time period in order to find the combination of parameters fitting each time period best. The best parameters, throughout the time periods, were all related either to a steeper river or a river limited in its ability to grow wider in times of high discharge. Therefore, to lower the erosion rates of the Elbe, measures must be taken that either enable the river to widen during high discharge e.g., the breaking of dams, or measures ensuring that the river flows less steep e.g., the reconnection of abandoned channels.

Zusammenfassung

In dieser Arbeit wurde integrated nested Laplace approximation (INLA) angewandt, um diejenigen Parameter zu identifizieren, die statistisch den größten Einfluss auf den deutschen Teil der Elbe haben. Dazu wurden das R Paket INLA von Rue et al. 2009 sowie ein Fork der Funktion `INLAsstep()` aus dem R Paket `INLAutils` (Redding et al. 2019) genutzt. Insgesamt wurden 67 Umweltparameter in ihrer Fähigkeit untersucht, die jährlichen Erosionsraten der Elbe für die drei Zeitscheiben 1898–1995, 1995–2004 und 2004–2017 zu beschreiben, um die, für die Erosion relevantesten, Parameter zu identifizieren. In der Kombination dieser Parameter wurden für jede der drei Zeitscheiben je 6017 Modellformeln getestet. Die besten Ergebnisse lieferten stets Parameter, die entweder einen Einfluss auf das Gefälle des Flusses hatten oder auf dessen Fähigkeit, sich bei steigenden Abflussbedingungen auszudehnen. Um der Erosionsproblematik der Elbe entgegenzuwirken, werden deshalb Maßnahmen empfohlen, die entweder dazu beitragen, dass die Elbe sich bei steigendem Pegel stärker ausdehnen kann, wie z. B. das Durchbrechen von Dämmen, oder Maßnahmen, die dafür sorgen, dass der Flusslauf der Elbe weniger steil verläuft, wie beispielsweise das Wiederanschießen abgeschnittener Altarme.

Table of Contents

Abstract	I
Zusammenfassung	I
1 Introduction	1
1.1 General Introduction	1
1.2 Aims and approaches	2
2 Study area	4
2.1 The River Elbe in general	4
2.2 Geology of the River Elbe	6
3 Material & Methods	7
3.1 Software	7
3.1.1 R & RStudio	7
3.1.2 R-INLA	8
3.1.3 GRASS GIS	9
3.1.4 ArcMap	10
3.2 The datasets	10
3.2.1 Tiling along River Elbe (<i>hydflood::spdf.tiles_elbe</i>)	10
3.2.2 Cross-section areas (<i>spdf.afe_csa</i>)	10
3.2.3 The axes	11
3.2.4 Hectometers along River Elbe (<i>spdf.hectometer</i>)	12
3.2.5 Water level data frame (<i>df.flys</i>)	12
3.2.6 Digital elevation models (DEM)	12
3.2.7 Potential natural vegetation (PNV)	12
3.2.8 Digital landscape model 1:25000 (DLM25)	13
3.2.9 Riverbed erosion rates	14
3.3 functions	14
3.3.1 getSlope()	14
3.3.2 getDEMmin(), getDEMmax()	15
3.3.3 Functions to query the percentage of inundated floodplain for FLYS3 water levels	16
3.3.4 getDiffbetweenMLQMHQ()	16
3.3.5 intoDB()	16
3.3.6 getAngle()	17
3.3.7 getAxisADiff/PDiff()	17
3.3.8 getCurve()	17

3.3.9 getDLM()	18
3.3.10 getRatio()	18
3.3.11 getVegType().....	18
3.3.12 INLAstep() for model creation.....	18
3.4 Model selection.....	19
4 Results	20
4.1 The erosion rates for the different time periods	20
4.2 Queried hydrogeomorphic variables	22
4.2.1 Minimum height of the DEM.....	22
4.2.2 Difference in river axes and river slope	23
4.2.4 Proportions of active floodplain between MLQ and MHQ	24
4.2.5 Height differences between FLYS3 water levels MLQ and MHQ	24
4.2.6 Proportion of PNV type hardwood floodplain wet (HFW)	25
4.2.7 Proportion of PNV type reed (RE).....	25
4.2.8 Proportion of PNV type periodically flooded (PF)	25
4.2.9 Proportion of DLM25 landuse type grassland	26
4.2.10 Proportion of DLM25 landuse type human	26
4.3 Best models	27
4.3.1 1898–1995.....	28
4.3.2 1995–2004	30
4.3.3 2004–2017	32
5 Discussion	33
5.1 Results of the query functions.....	35
5.2 Time period 1898–1995	36
5.3 Time periods 1995–2004 and 2004–2017	37
6 Conclusion.....	38
Literature	40
Appendix	43

Figures & Tables

Figure 1 Map of the floodplain of the study area	5
Figure 2 Comparison of historical and recent river axis	11
Figure 3 Depiction of <i>getSlope()</i>	14
Figure 4 A depiction of the general querying workflow	16
Figure 5 A depiction of <i>getAngle()</i>	17
Figure 6 A graphical explanation of the functions <i>getAxisADiff()</i> .and <i>getAxisPDiff()</i>	17
Figure 7 A depiction of <i>getRatio()</i>	17
Figure 8 A depiction of the cross-section areas.....	18
Figure 9 A depiction of <i>getRatio()</i> works.	18
Figure 10 The annual fairway erosion rates plotted against the kilometrage of the Elbe of the time periods 1898–1995 (A), 1995–2004 (B), and 2004–2017 (C).	21
Figure 11 The returns of the function <i>getDEMmin()</i>	22
Figure 12 The returns of the functions <i>getSlope()</i> and <i>getAxisADiff()</i>	23
Figure 13 The returns of the functions <i>getDiffbetweenMLQMHQ()</i> and <i>getPbetweenMLQMHQ()</i>	24
Figure 14 Proportions of the active floodplain of PNV type HFW..	25
Figure 15 Proportions of the active floodplain of PNV type RE.....	25
Figure 16 Proportions of the active floodplain of PNV type PF.	25
Figure 17 Proportions of the active floodplain of DLM25 type grassland	26
Figure 18 Proportions of the active floodplain of DLM25 type human a	26
Figure 19 Erosion rates for the three time periods 1898–1995, 1995–2004, 2004–2017	27
Figure 20 Comparison between fitted and measured erosion rates 1898–1995.	28
Figure 21 Frequency analysis 1898–1995.....	28
Figure 22 Comparison of the waic of the different model types 1898–1995.	29
Figure 23: Graphical summary of model m1.....	29
Figure 24 Graphical summary of model m2.....	30
Figure 25 Comparison between fitted and measured erosion rates 1995–2004.....	31
Figure 26 Frequency analysis 1995–2004.....	31
Figure 27 Comparison of the waic for the different model types 1995–2004.....	31
Figure 28 Comparison between fitted and measured erosion rates 2004–2017	32
Figure 29 Frequency analysis 2004–2017.....	32
Figure 30 Comparison of the waic for different model types 2004–2017	33
Figure 31 Graphical summary of model m3.....	33
 Table 1 Overview over the different potential natural vegetation types derived from annual inundation times.....	13
Table 2 Overview over the vegetation types and their respective coding within DLM25.....	13
Table 3 Overview over the query functions and the associated parameters derived from them.....	15

1 Introduction

1.1 General Introduction

Natural floodplains form a connected and interrelated ecosystem with the river (Meyer 2017). They are characterized by alternating floods and dry periods and are among the most biologically productive and diverse ecosystems on earth. Together with their rivers, they combine to form a complex, dynamic physical and biological system providing a variety of important environmental services (Witt 1993, Tockner & Stanford 2002). Apart from their obvious function as retention areas in times of high discharge, floodplains also play a role in water filtration and as drinking water extraction sites. They also improve the general water quality of rivers and serve as habitats to a great diversity of species in general and especially to the endangered riparian forests. In central Europe floodplains represent the ecosystems with the highest species and habitat diversity (BMU & BfN 2009). Due to their high adaptability to varying degrees of discharge floodplains may also play a vital role in buffering the effects of climate change (BMU & BfN 2009).

Natural floodplains are characterised by such a high biodiversity because they offer a high variety of habitats, essentially forming a mosaic of areas that develops due to many disturbances (Schindler et al. 2016, Ward 1998). These areas differ in age, humidity, sediment properties, productivity, diversity, and composition and succession state of biota. They are also characterised by inhabiting a multiplicity of terrestrial and aquatic generalist and specialist species (Schindler et al. 2016). It is therefore apparent that floodplains have a strong claim to being protected.

Besides being among the most biologically diverse ecosystems on earth, floodplains are also among the most threatened ecosystems. They are suffering from habitat alteration, cultivation, flood and flow control measures, species invasion, and pollution. In Europe and North America up to 90 % of floodplains are already functionally extinct and in every part of the world the remaining natural floodplains are disappearing at an accelerating rate (Tockner & Stanford 2002). Increasing erosion rates that may come from river straightening, the employment of spur dikes and various environmental factors also threaten floodplains to a great degree. As the river erodes deeper into the ground, the water level compared to sea-level subsides and the level of the ground water surface sinks (Meyer 2017). With this sinking water level, a larger amount of additional water is needed to inundate the floodplains. Therefore, the floodplain gets drier the further the erosion proceeds, resulting in changes of the floodplain ecosystem and ultimately a loss of biodiversity.

1.2 Aims and approaches

Sustainable conservation of floodplains implies knowledge of the basic geomorphological and ecological processes that interplay at the landscape scale (Bornette et al. 1998). The geomorphic features of floodplains reflect complex interactions between climate, catchment geology, topographic relief and fluvial dynamics (Ward et al. 2002). This work aims at closing some of the gaps in our knowledge and understanding of the incision processes of rivers and to give some advice at how to prevent further incision and coupled losses of floodplains and biodiversity. Also, the work aims at creating new information by connecting existing datasets and creates means to identify the parameters that have the strongest impact on a river's erosion rate. All the datasets used to draw out the analyses within this work are spatial datasets and all the parameters derived from these datasets are spatially distinct parameters.

The main obstacle to overcome when working with such spatial data, within the classical statistical framework, is spatial autocorrelation. A correlation in general measures the strength of a statistical relationship between two variables. A positive correlation means that an increase in one of the two variables is associated with an increase in the second variable and a decrease in one is associated with a decrease in the other variable. A negative correlation means that one variable decreases when the other increases and vice versa. When none of these relationships is apparent between two variables, they are considered uncorrelated or independent (Fahrmeir et al. 2016). Autocorrelation, on the other hand, means that one variable correlates with itself. The easiest type of autocorrelation to understand is temporal autocorrelation. For instance, if the weight of a patient gets measured every day, it makes sense that the weight at a given day is similar to the weight of the day before and the day after. Or in general that observations that are close to each other in time are also similar in measurement. It is unlikely that the weight differs multiple kilograms between two days, but it is not unlikely that it does so over the period of years. Therefore, observations that are not close to each other in time may differ significantly. Analogously, autocorrelation can be caused by spatial rather than temporal proximity. Observations that are close to each other spatially tend to be close to each other in measurement values (Hijmans 2021).

To overcome this issue of spatial autocorrelation the statistical analyses within this work employ the Bayesian framework rather than classical statistics. Bayes' theorem is a prescriptive statistical framework that is based on the combination of probabilities. It differs from classical statistics by predicting posterior probabilities not only by the data collected within a certain experiment, but also by including the prior probabilities of a phenomenon (Tschirk 2014). Therefore, the typical Bayesian workflow consist of three main steps: (I) capturing available knowledge about specific parameters in a statistical model via the prior distribution, (II)

determining the likelihood function via the data gained during the experiment, and (III) combining both the likelihood function and the prior distribution to form the posterior distribution. The posterior distribution then holds the updated knowledge about a phenomenon, balancing the prior knowledge with observed data (van de Schoot et al. 2021).

To study data prone to spatial autocorrelation, the Bayesian framework has been used widely in recent years across a variety of research fields. Mitra (2009) used it to analyse intersections in the US prone to injury crashes (Mitra 2009). Chun, Kim & Campbell (2012) used Bayesian methods to control for spatial autocorrelation in environmental justice research, analysing how toxics released into the environment may concern some ethnics more than others within the Maricopa County in California (Chun, Kim & Campbell 2012). Bayesian models are also widely used in ecological studies. March et al. (2013) for example, used a Bayesian spatial approach to predict the occurrence of different species of seagrass to aid in seagrass management and the assessment of the ecological status of seagrass meadows (March et al. 2013). The usage of Bayesian methods within ecological sciences got particularly popular since the publication of INLA and the R package *R-INLA* providing easier access to spatial Bayesian inference models (Rue et al. 2009). Within this work, *R-INLA* was used to statistically analyse the available spatial data. INLA is short for integrated nested Laplace approximation. It is a method for approximating Bayesian inference and is used here to make spatial predictions (R-INLA 2021). The R package *R-INLA* is a further development of the stand-alone software *INLA* (Rue, Martino & Chopin 2009, Martins et al. 2012) and builds the gateway between this INLA and *RStudio* or the R programming language in general (Turkman et al. 2019).

Spatial modelling is an important statistical field that is computationally usually very challenging. In most well-established methods for spatial modelling the tool for capturing spatial dependency is the Gaussian random field, which is hard to use when – like in this work – there is a lot of data. One workaround for this problem is the usage of the SPDE approach (stochastic partial differential equations approach) that is used by the INLA method within the R package *R-INLA* (Rue, Martino & Chopin 2009, Lindgren, Rue & Lindstrom 2011). The package offers a computationally efficient implementation of the SPDE approach, as well as a relatively simple interface for model adjustments (Bakka et al. 2018) making it a vital tool to overcome the statistical issues discussed above. The usage of the package is discussed in further detail within section 3.1.2. In short, the function *inla()* within the R package *R-INLA* gets an input of a set of spatially located data observed at discrete points and information about the spatial relationships between these datum points, which is key to account for spatial autocorrelation. The function then uses this spatial information to model a spatial phenomenon based on a variable number of input parameters thought to analyse this phenomenon. Here, it was used to relate

the erosion rates of the River Elbe for three different time periods to a variety of environmental parameters.

R-INLA was employed to answer three main research questions: (I) Can a complex problem like a river's erosion rate be simplified to few environmental parameters?, (II) which parameters are the most relevant in terms of explaining a river's erosion?, and (III) what do the findings of the *R-INLA* models imply for applied river conservation measures? Or in other words, what ought to be done in the future to get erosion rates and the corresponding problems under control?

2 Study area

2.1 The River Elbe in general

The River Elbe rises in the Krkonoše Mountains in the Czech Republic, has a length of 1094 km, and flows into the North Sea at Cuxhaven in Germany (Fig. 1). In total, it drains the fourth largest catchment in Europe with 148268 km², has a mean discharge of 861 m³/s (FGG Elbe 2021) and a length of 727 km within Germany (FGG Elbe 2020). The river's shape, as it is today, was influenced to a great degree by settlement, cultivation, and energy generation purposes as well as by flood prevention measures and measures to improve shipping like straightening, groyne construction, and embankment (Nestmann & Büchele 2002). These anthropogenic measures severely impacted the floodplain, which has lost between 50 and 90 % of its original area. Nevertheless, compared to other German federal waterways the damage to the natural floodplain of the river Elbe is relatively small (BMU & BfN 2009) and the Middle Elbe Biosphere Reserve is home to one of the last connected stretches of riparian forests in central Europe (WWF Deutschland 2021).

The Elbe is characterised by dynamically changing flow rates and coupled aggradation and erosion processes that occur naturally. These short-term changes in discharge are interlinked with a high variability of water on the adjacent floodplains and therefore a highly instable system that, as discussed above, results in a high level of biodiversity (Nestmann & Büchele 2002). On the other hand, the Elbe also undergoes some long-term changes in fluvial morphology that may be linked to the anthropogenic altering of the waterway. Riverbed incision is a known issue for certain parts of the river. It means the water level of the main river progressively deepens which leads to a progressive isolation of the river from its floodplain and consequently decreases the flood frequency and seepage supply to the riverine wetlands (Borrette et al. 1998).

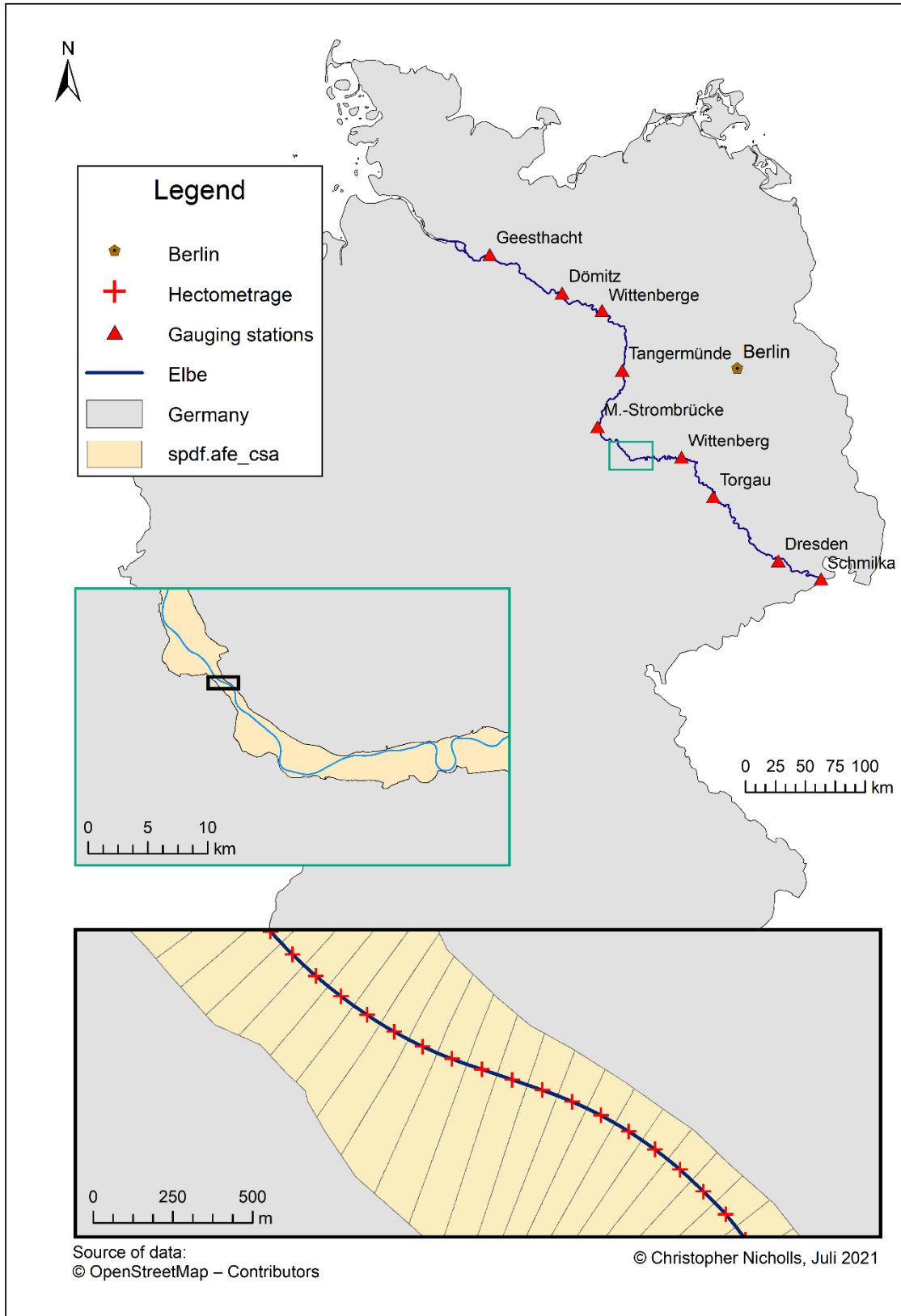


Figure 1 Map of the floodplain of the River Elbe between Schmilka and Geesthacht, Germany. Illustrated, in zoomed panels, are the river axis (*sldf.axis*), the river stationing (*spdf.hectometer*), and a segmented floodplain (*spdf.afe_csa*), which builds up on input data of a 1D hydraulic model (SOBEK). These datasets form the basis of the applied spatial model and are described in detail in section 3.2.

2.2 Geology of the River Elbe

For the development of the German part of the River Elbe the most important geological time periods were the Quaternary and the Holocene. These were characterised by alternating warm and cold periods in which the geomorphological processes were mainly powered by either glacier formation and movement or melting water. The sediments that were deposited in northern Germany during the Elster, Saale, and Weichsel cold periods and during the Holocene represent the material the Elbe had to cut itself into and ultimately formed the course of the river as seen today (Saucke & Brauns 2002). Over time the incision of the Elbe uncovered three prominent erosion barriers in Torgau and Magdeburg. The Torgauer Felsen at Torgau is located around river-km 155. Geologically it is a rhyolite and resulted from the Variscan orogeny. The erosion barriers in Magdeburg are the Domfelsen and the Herrenkrugfelsen. They lie around river-km 325 and 329 respectively and geologically are conglomerates of the Rotliegendes and carbonic greywacke (Faulhaber 1996, Saucke, Rommel & Brauns 1999, Simon 2018, Saucke & Brauns 2002).

The beginning of the Elbe within Germany, however, is characterised by a different geology. When crossing the Czech/German border at Schmilka the river flows through the Elbe Sandstone Mountains, the German part of which is called Saxon Switzerland. Here, the river is flanked by high and steep rock walls that arose during the first half of the Late Cretaceous and are made up by Cretaceous sedimentary rocks. The Elbe leaves the Saxon Switzerland around Dresden (Fig. 1). Further downstream the glacial influence became predominant in shaping the landscape, as discussed above (Mannsfeld & Bernhardt 2008, Saucke & Brauns 2002).

Another important geomorphological feature worth discussing within this section is the composition of the riverbed. Its structure has implications on many characteristics of a river e.g., its hydraulics, ecological characteristics concerning benthos, or the sediment regime. The sediment regime in turn is an important factor when it comes to erosion processes as discussed within this work (Haunschild 1996, Saucke & Brauns 2002). The sediment regime of the upper course of the Elbe is mainly characterised by rubble until ca. river-km 150. Here, just before the erosion barrier Torgauer Felsen the main size of the sediment decreases rapidly from rubble to medium gravel and then after the erosion barrier from around 15 mm in diameter continuously to small gravel with 3 mm in diameter. The sediment regime then reaches a highly variable state from river-km 300 to river-km 350 around the erosion barriers of Magdeburg. Here, the main grain size alternates between 3 and 12 mm. Afterwards in the lower course of the river, the grain size continuously decreases. The main grain size from river-km 400 onwards is sand with a mean diameter of around 1 mm (Haunschild 1996).

3 Material & Methods

The general idea of this work was to characterise the Elbe with a multitude of environmental or landscape parameters and to check which of these parameters or combination of parameters has the greatest influence on the erosion rate of the riverbed. To achieve this, the R package *elbe1d* was programmed. Together with the usage of the geographical information systems GRASS GIS and ArcMap, that were used to organize and process the data, 16 functions that derive a total of 67 parameters were written (Tab. 3). The ideas behind the functions are briefly discussed within the next section. To get a full understanding of the code, I refer to the functions' documentations within the R package hosted on GitLab (<https://git.aqualogy.de/chris/elbe1d>).

After presenting the software used within this work, the query functions and datasets are described, and insights into the application of the *R INLA* package and a forked version of the *INLAstep()* function, providing model selection facilities, are given.

3.1 Software

This section discusses the software used within this work. It gives details about versions, download sites and some hints as to why these programs were used.

3.1.1 R & RStudio

R is both a programming language and an open-source software for statistical data analysis. In this work it was used within the programming environment RStudio. R was used in version 4.0.3, the RStudio version used was 1.3.1093. Apart from being an open-source software, R offers a lot of additional advantages. It is available for Windows-, Mac- and Linux-systems, has gateways to a many other applications like GRASS GIS or LaTeX, offers an easy environment to couple multiple, reproducible analysis steps and has a great toolbox for data visualization. The greatest advantage of R, however, is its variability based on its package structure. Every user is free to develop and contribute own packages with own functionalities that might be useful for other users. Many of them are available at CRAN (Comprehensive R Archive Network), R's central software repository (Schmidt & Boyd 2020, Mizumoto 2016) and many more – like the package developed in this work – are available on version control sites like GitHub or GitLab.

Base R is simply a command-line tool with few features, but the usage of RStudio and additional packages gives R great power. RStudio makes programming in R much easier because it offers an intuitive user interface in which to program (Schmidt & Boyd 2020). It is an integrated development environment (IDE) available for Windows, Mac and Linux users that includes a console, a syntax-highlighting editor with direct code execution, as well as tools for

plotting, history, debugging, and workspace management. It also offers a direct gateway to GitLab or other git-based version control systems (RStudio Team 2021), and built-in tools to create markdown documents. The coupling with version control systems enables multiple users to aid in cooperative software development and was an indispensable asset during the development of the package *elbe1d*.

3.1.2 R-INLA

As discussed within Chapter 1, INLA was integrated into this study via the R package *R-INLA* (Rue, Martino & Chopin 2009, Lindgren, Rue & Lindstrom 2011). The stable version – that was used within this work – as well as the testing version of the package can be downloaded via the package website (<https://www.r-inla.org/download-install>). Under the hood *R-INLA* utilised the *PARDISO 7.2 Solver* for thread-safe, high-performance, robust, memory efficient and easy solving of large sparse symmetric and unsymmetric linear systems of equations on shared-memory and distributed-memory multiprocessor systems to accelerate computations on a HPC (Schenk 2021, van Niekerk et al. 2019).

To use INLA for model fitting, a formula object must be created, holding the linear predictor i.e., the set of variables that ought to predict the predicted variable (Moraga 2019). The *inla()* function then takes this formula together with the specified parameters “*family*” – describing the type of distribution the predicted data presumably follows, “*data*” – consisting of an *inla.data.stack* object containing all needed data, and an *spde* model object (*SPDE*) – giving the opportunity to include spatial relationships between the data into the model. The *inla.data.stack* object in turn consists of four parts. These are “*data*” – a list with data vectors, “*A*” – a list of projector matrices, “*effects*” – a list of effects or predictors (i.e., covariates), and “*tag*” – a character with a label for this group of data. The *SPDE* is created from a one- or more-dimensional mesh between the points in which the data is located. Via the projector matrix *A*, the projection of the *SPDE* gets mapped to the observed points (Gómez-Rubio 2020) i.e., the spatial locations on the mesh are translated into vectors in the model (Albery 2021). Within this work, a one-dimensional mesh was created along the hectometer points of the River Elbe.

When creating the mesh, not only the spatial relationships between the datum points are set but also the precision of the model gets defined to some extent. Within two-dimensional models this is done by setting the triangle size of the mesh, determining how precisely the equations will be tailored to the data. Smaller triangles increase precision but also exponentially increase the required computing power (Albery 2021). Within the one-dimensional models presented here, the mesh is defined as a line through the hectometer points and, similarly to triangle sizes over a two-dimensional mesh, the precision is regulated by the number of nodes set

along this one-dimensional mesh. Here, the distance between two nodes was set to 1000 m i.e., one node for every ten hectometer points along the river axis.

As discussed above, the *inla()* function works with spatial data and information about the spatial relationships between the datum points. On the other hand, *inla()* can also be used without incorporating spatial relationships. It can either be run just using fixed effects i.e., considering the variables specified in the formula, or it can be run using fixed effects coupled with random effects. The version that accounts for spatial autocorrelation is using fixed and random effects as well as spatial relationships all together. Within this work all three approaches were tested to check whether spatial autocorrelation indeed has a big influence on the data.

The *inla()* function returns an object with information about the fitted model, the posterior marginals of the parameters, the linear predictors, and the fitted values. It also returns different estimates of how good the model fits the predicted data (Moraga 2019). To evaluate the models tested within this work, the Watanabe-Akaike information criterion (waic) was used, with smaller values suggesting a better fit. The *family* parameter was set to “gaussian”, and the covariates were the parameters derived from the query functions discussed above. For the creation of the one-dimensional mesh and the *SPDE*, the datum points within the dataset *spdf.hectometer* were used. The predicted variables were the erosion rates from the three time periods 1898–1995, 1995–2004, and 2004–2017.

3.1.3 GRASS GIS

GRASS is short for “Geographic Resources Analysis Support System”. It is an open-source geographical information system (GIS) to process geographical data both in raster and vector format. GRASS GIS offers more than 300 modules to process raster, vector, and point data, has gateways to a variety of database management systems, can work with a large amount of different data types including ESRI shapefiles (Neteler 2001), and is easily integrated into the R environment via the R package *rgrass7*. This package provides facilities to use all GRASS commands from the R command line. In this work the package version 0.2-5 was used (Bivand 2021). GRASS GIS was used in version 7.8 downloaded via the official website distributed by OSGeo4W.

GRASS GIS is structured along three consecutive layers. The first layer is the database. It is the directory in which ultimately all the GRASS GIS data is stored. It contains one or multiple locations. The location is the second layer of the GRASS GIS structure and gives the framework for all the data it contains by providing a unique coordinate reference system, map projection, and geographical boundaries. Within the location several mapsets as well as permanent data can be stored. The mapset represents the actual working environment. It holds the

vector and raster data that are used by the functions within the package (GRASS Development Team). All the operations done within this work were drawn out within the GRASS GIS location *ELBE_Binnen*. It contains 49 mapsets that represent the 49 tiles along the river (see Section 3.2.6). Every tile is thereby described by a specific set of datasets like data describing the landuse, cross-section areas, the part of the floodplain inundated during different discharge scenarios, the types of waterbodies within the tile, or different potential natural vegetation types. All these datasets are named the same within the different mapsets which allows for functions implemented in R to open different mapsets and execute the same operations without the need to adjust file names.

3.1.4 ArcMap

ArcMap is a commercial GIS software that comes within the ArcGIS package produced by the ESRI company (ESRI 2020). Within this work the version 10.3.1 was used that is provided by the University of Koblenz via license server. Its intuitive user interface and a vast variety of tools are its main advantages. Here, ArcMap was mainly used to prepare, combine, and create datasets that were afterwards used within the combined environment of GRASS GIS and R Studio.

3.2 The datasets

The following paragraphs present the datasets that come with the package *elbe1d* and are essential for it to function. The origins as well as possible manipulations of the data are drawn out.

3.2.1 Tiling along River Elbe (*hydflood::spdf.tiles_elbe*)

This dataset of type *SpatialPolygonsDataFrame* is derived from the R package *hydflood* (Weber & Rosenzweig 2020). It contains 49 rectangular polygons set along the German interior parts of the Elbe i.e., from the Czech/German border to the barrage of Geesthacht, that are from now on referred to as tiles. These tiles split the Elbe as well as its active floodplain into 49 separate regions that form the units described by the 49 GRASS GIS mapsets discussed within section 3.1.3. They also form the computational units in which all the spatial queries for the predictor variables took place and the DEM has been provided.

3.2.2 Cross-section areas (*spdf.afe_csa*)

This dataset is of type *SpatialPolygonsDataFrame* and represents the central spatial object within the package. It was derived from the dataset *spdf.active_floodplain_elbe_csa* from the R package *hydflood* that originally contained a polygon of the active floodplain between the dikes along the German interior parts of the Elbe. This polygon was originally produced for the floodplain status report (BMU & BfN 2009) at a scale of 1:25000 and later updated with recent

flood protection measures and manually improved with recent digital elevation models and aerial images at a scale < 1:10000 (Weber & Rosenzweig 2020). The abbreviation *afe_csa* is short for “active floodplain Elbe cross-section areas”. The dataset contains 5850 features representing the hectometers (section 3.2.4) along the axis. These features are small rectangular polygons subdividing the 49 large tiles of the dataset *spdf.tiles_elbe* into smaller cross-section areas (Fig. 1). These cross-section areas couple the information returned by the functions to discretely located areas along the river axis. This allows the data querying in a resolution of 100 m for the whole river and is also the requirement to account for spatial autocorrelation within the statistical analysis.

3.2.3 The axes

To compare and to calculate differences between the present-day Elbe and the Elbe around the year 1830 the recent axis of the river as well as the historical one are needed. The axis of the Elbe as it is today, is stored within the dataset *sldf.axis.rda*. The historical axis is stored in *sldf.axis1800.rda*.

3.2.3.1 Recent axis (*sldf.axis*)

This dataset of type *SpatialLinesDataFrame* was derived from the open-source dataset VerNet-BWaStr of the German Federal Waterway Network (Verkehrsnetz Bundeswasserstraßen) (Fig. 2). It is the official dataset published by the German Federal Waterways and Shipping Administration (WSV 2020) containing all federal waterways of Germany.

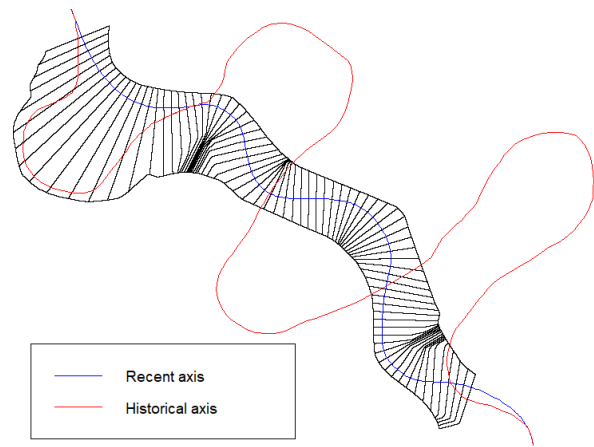


Figure 2 Present (*sldf.axis*) and historical (*sldf.axis1800*) axis between river kilometer 161 and 169.

3.2.3.2 Historical axis (*sldf.axis1800*)

This dataset of type *SpatialLinesDataFrame* contains the historical axis of the river dating to around 1830 (Fig. 2). It has been reconstructed from maps dating around this time for the German part of the Elbe for river-km 0–585. The reconstruction was done by tracing the former course of the river from historical maps and afterwards scanning and digitalising the resulting copies. The full course was reconstructed using multiple maps with different scales. To merge them, the scale was set to 1:200000 (Rommel 2000).

3.2.3.3 Axis buffer (*sides_axis_buffer*)

This dataset contains the recent axis of the river Elbe combined with the extent of the active floodplain. The dataset essentially consists of four polygons. Two of these polygons represent buffers of 50 m on both the left and right-hand side of the river axis and the other two contain

the parts of the active floodplain that are not covered by the buffer areas, again split in the parts on the left and the right-hand side of the river. Its main purpose is to reduce the computing time of the function *getRatio()* presented in section 3.3.10.

3.2.4 Hectometers along River Elbe (*spdf.hectometer*)

This dataset of type *SpatialPointsDataFrame* contains the official hectometer points along the axis (*sldf.axis*) of the Elbe. It was also derived from the open-source dataset VerkNet-BWaStr (WSV 2020). The hectometer points represent the official stationing used to setup the SOBEK models for *df.flys* and the cross-section areas stored in *spdf.afe_csa*. Their stationing (*station_int*) is the basis to join the various datasets and their distance to the origin of *sldf.axis* on the Czech border (*station_true*) is used to place the model data on the one-dimensional mesh.

3.2.5 Water level data frame (*df.flys*)

This *data.frame* contains the water levels for the different discharge scenarios queried within the functions that calculate the percentage of inundated floodplain. It contains 30 stationary 1D water levels for the rivers Elbe and Rhein and is part of the R package *hyd1d*. These water levels have been computed by means of the 1D hydraulic model SOBEK and are available for the full length of the Elbe with a resolution of 200m along the official river stationing (Weber & Hatz 2020). The *data.frame* is not directly included into the R package *elbe1d* but gets queried indirectly by calling the functions provided by the R package *hyd1d*.

3.2.6 Digital elevation models (DEM)

The used DEM is an open-source dataset (Weber 2020) and contains 49 raster datasets (tiles) of floodplain elevations with 1 m spatial resolution outlined by *hydflood::spdf.tiles_elbe*. Elevation data were derived from laser scanning for terrestrial parts and in permanently inundated parts of the floodplain around the fairway by echo sounding between 2005 and 2011. Permanently inundated parts of the floodplain outside the fairway are included with their water surface making these areas error prone.

3.2.7 Potential natural vegetation (PNV)

The distribution of potential natural vegetation has been provided by the BfG. It is a reclassification of the mean of annual flood durations between 1990 and 2019 computed with *flood3* according to Tab. 1 on the basis of the used DEM (Weber 2020). It has 1 m spatial resolution and was used by the function *getVegType()* to query the area covered with different types of potential natural vegetation for a specified cross-section area.

Table 1 Overview over the different potential natural vegetation types derived from annual inundation times.

Flood duration (days/year)	Class	Type of vegetation	Abbreviation
$0 \Rightarrow x < 5$	1	Zonal forest	ZF
$5 \Rightarrow x < 50$	2	Hardwood floodplain, dry	HFD
$50 \Rightarrow x < 80$	3	Hardwood floodplain, wet	HFW
$80 \Rightarrow x < 140$	4	Willow forest	WF
$140 \Rightarrow x < 160$	5	Willow shrubbery	WS
$160 \Rightarrow x < 220$	6	Reed	RE
$220 \Rightarrow x < 300$	7	Periodically flooded	PF
$300 \Rightarrow x \leq 366$	8	Water	WA

3.2.8 Digital landscape model 1:25000 (DLM25)

Table 2 Overview over the vegetation types and their respective coding within DLM25 (GeoBasis-DE/BKG 2021)

DLM 25 attributes objartint, vegint	Class	Type of vegetation
1010 ,1011, 1012, 1014	1	agriculture
1020 ,1021, 1022	2	grassland
1030, 1031, 1040, 1050, 1051, 1052, 1060	3	gardening, vine, plantations
1021, 1100	4	deciduous forest
1022, 1200	5	coniferous forest
1023, 1300	6	mixed forest
43003	7	wood
43004	8	heath
43005	9	bog
43006	10	swamp
43007	11	free
144001, 44002	12	flowing water
44003, 44004, 44005, 44006, 44007	13	standing water

The DLM25 provided by the Federal Agency for Cartography and Geodesy contains landscape information at scale 1:25000 in vector format (GeoBasis-DE/BKG 2021). It was imported in GrassGIS, transformed, and natural land cover types (Tab. 2) were rasterized with 1 m spatial resolution. It gets used by the function *getLanduse()* to query the area covered with different types of landuse for a specified cross-section area.

3.2.9 Riverbed erosion rates

To derive riverbed erosion rates the BfG provided riverbed heights for four times (1898, 1995, 2004, and 2017). The riverbed heights for 1898 were derived from the “Elbstromwerk” (Elbstrombauverwaltung 1898). All later heights were recorded by echo sounding by the waterway and navigation authorities and spatially queried as mean riverbed height of the fairway within each cross-section area. Data for 1995 were recorded between 1993 and 1999, for 2004 between 2003 and 2004, and for 2017 in the years 2016 and 2017 through more or less continuous routine echo sounding campaigns. The four riverbed heights were recomputed to erosion rates by dividing the change in riverbed height between the time periods through the number of years in between the measurements.

3.3 functions

This section presents the functions within the R package that were used to query the data and produce explanatory variables for the prediction of the riverbed erosion rates. In sum 67 numerical parameters describing the Elbe in a resolution of one hectometer were derived and the results collected in a database. In addition to the query functions, the package contains functions to aid in data visualization and the automatic creation of the database as well as a developmental version of the function *INLAsstep()* from the R package *INLAutils* (Redding et al. 2017) which is presented within section 3.3.12.

3.3.1 getSlope()

getSlope() calculates the average slope at a given hectometer point along the axis of the river Elbe. To calculate the slope between two points, the distance between the points and the difference in their heights of the *FLYS3* water level for MQ is needed (Fig. 3). Here, the distance along the axis of the river Elbe and the difference in height between the mean water level above sea level at the starting and at the end point are used.

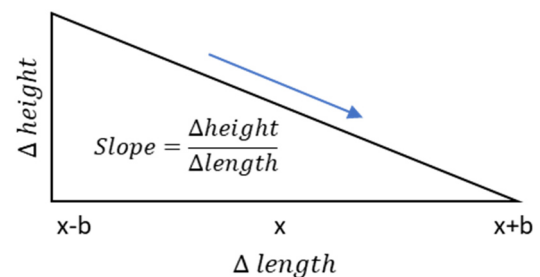


Figure 3 Depiction of the gradient triangle that symbolises the general functionality of the function *getSlope()*

By dividing the distance by the difference in height, the average slope is returned. The function takes additional parameters to dynamically calibrate the desired longitudinal buffer around a specified hectometer point for which the slope should be returned.

Table 3 Overview over the query functions and the associated parameters derived from them.

function (16)	Output (67)	Explanation
<i>getAngle()</i>	Angle	Steepness and direction of river curves.
<i>getAxisADiff()</i>	AxisADiff_b200, AxisADiff_b400, AxisADiff_b600, AxisADiff_b800, AxisADiff_b1000, AxisADiff_b2000, AxisADiff_b4000, AxisADiff_b6000, AxisADiff_b8000	Absolute difference in length between the historical and recent river axis for a specifiable area.
<i>getAxisPDiff()</i>	AxisPDiff_b200, AxisPDiff_b400, AxisPDiff_b600, AxisPDiff_b800, AxisPDiff_b1000, AxisPDiff_b2000, AxisPDiff_b4000, AxisPDiff_b6000, AxisPDiff_b8000	Relative difference in length between the historical and recent river axis for a specifiable area.
<i>getCurve()</i>	Curve_b200, Curve_b400, Curve_b600, Curve_b800, Curve_b1000, Curve_b2000, Curve_b3000, Curve_b4000, Curve_b5000	Steepness and direction of river curves for a specifiable area.
<i>getDEMmin()</i>	DEMmin	Minimum height of the DEM
<i>getDEMmax()</i>	DEMmax	Maximum height of the DEM
<i>getDLM()</i>	DLM_agriculture, DLM_bog, DLM_coniferous.forest, DLM_deciduous.forest, DLM_floating.water, DLM_free, DLM_gardening.vine.plantations, DLM_grassland, DLM_heath, DLM_mixed.forest, DLM_standing.water, DLM_swamp, DLM_wood	Area for different types of landuse
<i>getDiffbetweenMLQMHQ()</i>	DiffbetweenMLQMHQ	Difference in water level between mean low and mean high discharge
<i>getMaxFlowLength()</i>	MaxFlowLength	Maximum flow path length
<i>getPbelowMHQ()</i>	PbelowMHQ	Percentage covered at mean high discharge
<i>getPbelowMLQ()</i>	PbelowMLQ	Percentage at mean low discharge
<i>getPbelowMQ()</i>	PbelowMQ	Percentage at mean discharge
<i>getPbetweenMLQMHQ()</i>	PbetweenMLQMHQ	Difference in the percentage of the inundated active floodplain between mean low and mean high discharge
<i>getRatio()</i>	Ratio	Where within the floodplain the river flows (left/right)
<i>getSlope()</i>	Slope_b200, Slope_b400, Slope_b600, Slope_b800, Slope_b1000, Slope_b2000, Slope_b3000, Slope_b4000, Slope_b5000	The average slope over a specifiable area
<i>getVegType()</i>	VegType_HFD, VegType_HFW, VegType_PF, VegType_RE, VegType_WA, VegType_WF, VegType_WS, VegType_ZF	The area of different potential natural vegetation types derived from inundation times

3.3.2 *getDEMmin()*, *getDEMmax()*

The functions *getDEMmin()* and *getDEMmax()* determine the minimum and maximum height of a specific cross-section area within the tiles from the dataframe *spdf.afe_csa*. The functions use the digital elevation models (DEM) that exist for every tile along the river Elbe and contain spatially located information about the height of the terrain covered. These tiles are subdivided

in the cross-section areas evaluated by the functions (Fig. 4). They analyse every datum point within the DEM that falls into the area of a specified cross-section area, extract their heights, and return the minimum or maximum height detected.

3.3.3 Functions to query the percentage of inundated floodplain for FLYS3 water levels

The package contains different functions that calculate the percentage of the active floodplain covered by water at a given cross-section area along the axis of the Elbe at a certain discharge scenario. The functions check the DEM of a specified area and evaluate the percentage that lies lower than a certain water level. The functions *getPbelowMQ()*, *getPbelowMQ()*, *getPbelowMHQ()*, and *getPbetweenMLQMHQ()* query the proportion that is inundated at mean low discharge (MLQ), mean discharge (MQ), mean high discharge (MHQ), and the percentage of floodplain that is additionally inundated when the discharge scenario changes from MLQ to MHQ.

3.3.4 *getDiffbetweenMLQMHQ()*

getDiffbetweenMLQMHQ() calculates the difference in the water levels at MLQ and MHQ at a specific cross-section area. If the function returns a large value, it hints to a badly connected floodplain.

3.3.5 *intoDB()*

intoDB() manages the recording of queried data into the database *elbe1d*. It defines the functions that can write into the database and determines where into the database a new entry gets written. It also checks whether an entry has already been made and whether it needs updating. As a result, the database contains the return values of the functions for every hectometer i.e., every cross-section area. The database allows a comprehensive overview over the data created and aids in further work in form of cross-correlations and data visualization.

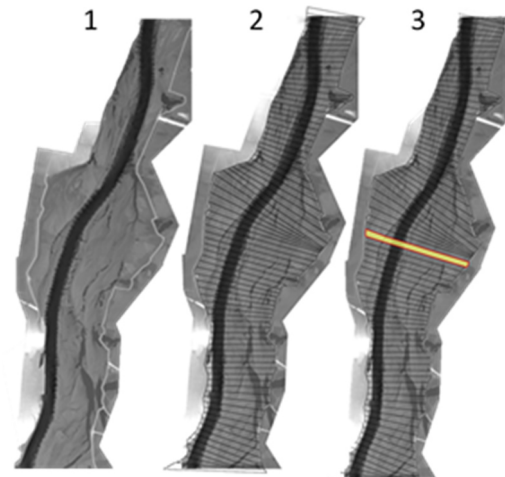


Figure 4 A depiction of the workflow the functions *getDEMmin()* and *getDEMmax()* as well as all of the other functions querying from raster data perform. (1) The specified tile, (2) the cross-section areas derived from *spdf.afe_csa*, and (3) the specified hectometer (yellow) the function ought to analyse.

3.3.6 getAngle()

getAngle() determines the steepness and directions of river curves. It uses the coordinates of three consecutive hectometer points along the river axis and determines the inner angle of the resulting triangle adjacent to the middle point (Fig. 5).

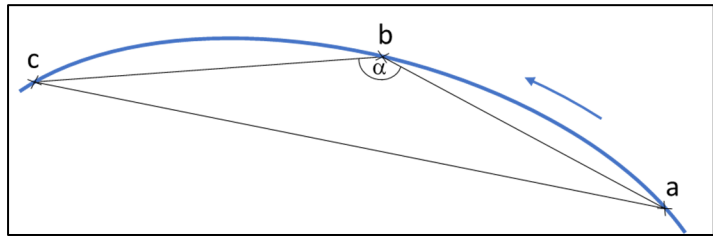


Figure 5 A depiction to aid in the understanding of the description of the functionality of *getAngle()*.

The direction is derived from the coordinates. If *c* is farther north than *a* and *c* is farther east than *b* or if *c* is farther south than *a* and *c* is farther west than *b*, the river takes a right turn. In any other case the river takes a left turn. The steepness of the curves is proportional to the calculated angle.

3.3.7 getAxisADiff/PDiff()

The function *getAxisADiff()* determines the absolute difference in length between the historical river axis and the recent course of the river (Fig. 6) for a specified hectometer within an adjustable buffer distance. The returned value is the difference in meters. A positive value indicates that the historical axis shortened over time, a negative value indicates that the recent axis is longer than the historical one. *getAxisPDiff()* is the same function but returns the percentage of the difference.

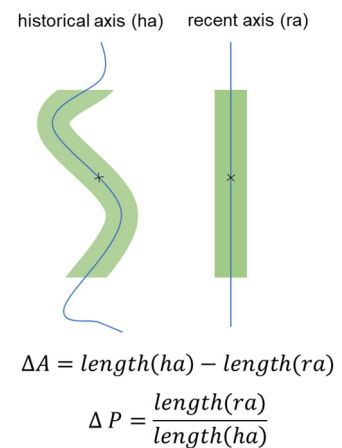


Figure 6 A graphical explanation of the functions *getAxisADiff()* and *getAxisPDiff()*

3.3.8 getCurve()

The function *getCurve()* determines whether the river follows a bend within a variable longitudinal buffer around a specified hectometer and returns a value describing the strength and direction of this bend. This is achieved by constructing a buffer of 50 m around the river axis and afterwards comparing the areas of the buffer on the left and right-hand side of the river (Fig. 7). Since the buffer area on the outer part of the curve is larger, the division of the two areas quantifies both the direction of the curve as well as the strength of the bend. The greater the difference in size between the inner and outer buffer areas, the steeper the river curve.

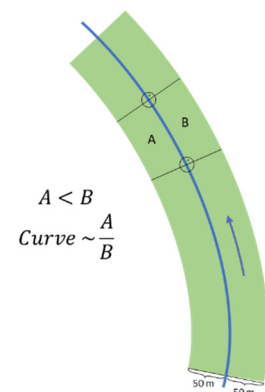


Figure 7 A depiction to aid in the understanding of the described functionality of *getRatio()*.

3.3.9 getDLM()

The function `getDLM()` takes in two parameters. Parameter `x` specifies the cross-section area and parameter `vt` specifies one or multiple types of landuse for which the information should be provided. The user can choose between the 13 types of landuse specified within the raster dataset `DLM_25_LANDUSE_RECLASS` (Tab. 2). This raster dataset then gets clipped to the specified area (Fig. 8) and the percentage of the specified vegetation type(s) within the specified area get(s) calculated and returned.

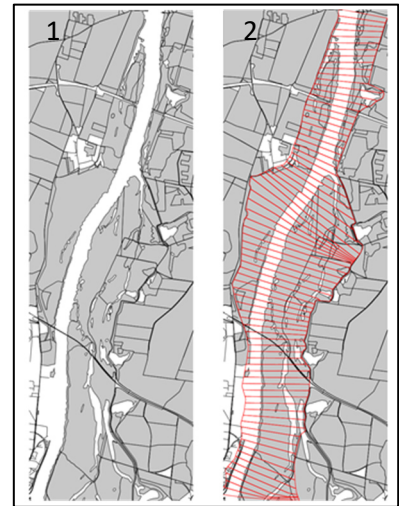


Figure 8 A depiction of the cross-section areas of the tile `e033_TANGERMUENDE` on top of the map containing the landuse data. The underlying data for a specified cross-section area gets queried by the function (GeoBasis-DE/BKG 2021).

3.3.10 getRatio()

`getRatio()` returns the position of the river within the active floodplain. It divides the area of the active floodplain on the left by the corresponding area on the right-hand side of the river (Fig. 9). The smaller the result the further left the river flows within the active floodplain at the specified cross-section area.

3.3.11 getVegType()

`getVegType()` returns the proportions of potential natural vegetation types for a specific cross-section area.

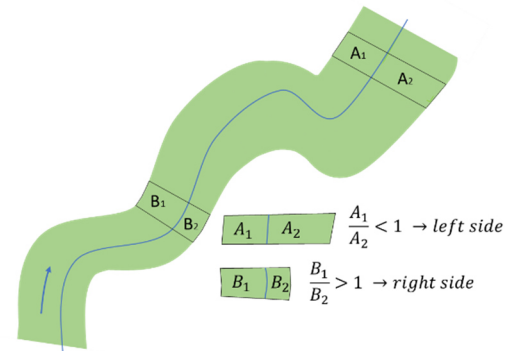


Figure 9 A depiction of the way the function `getRatio()` works.

3.3.12 INLAstep() for model creation

The function `INLAstep()` within the package `elbe1d` is a fork of the function `INLAstep()` from the R package `INLAutils` (Redding et al. 2017). It is essentially the same function with one added state to the input parameter “direction”. In the original function this parameter could only be set to either “forwards” or “backwards” stepwise model selection. The added third option “none” allows the integration of a loop into the function that is activated whenever the parameter is set thusly. `INLAstep()` itself, both the original and the development version, is a function that enables the user to test multiple functions with a prefixed set of variables in different combinations in their ability to predict values. In short it is a means to find the best model via the INLA method. The original `INLAstep()` has two different iteration methods to find the best fitting model. The “forwards” method takes only one of the given variables in the first round of testing and stores the one with the best result. In the second round of testing a second variable is added and the formula of two variables supersedes the previously stored formula if it fits the data better. In the third round a third variable is added

and so on until either every variable is within the formula, the fit of the tested formula does not improve, or some stop criterion is met. These stop criteria can be set via the input parameters “*thresh*” and “*include_max*”. *thresh* sets a threshold that must be exceeded for the iteration process to continue. If none of the untested formulas improves the stored function in an amount greater than the threshold, the function comes to an end. *include_max* sets a limit of variables to be included within the formulas to avoid functions that have a high quality of fit on paper but include a large number of variables, limiting their significance. The goodness of fit of the respective formulas is estimated by their waic. The second iteration method offered by the original *INLAstep()* function is the “backwards” method. Here the first round of iteration uses a formula containing all the variables. In the next step one variable is dropped from the formula and it is tested whether this improves the waic. In the next round another variable is dropped and so on until there is only one variable left or some stop-criterion is met. These two iteration methods have the advantage of finding a well-fitting formula without having to test all the possible variable combinations in order to save computing time. The added third state of the direction parameter „none“, however, tests all the possible variable combinations and stores the resulting formulas as well as the corresponding waic values in a *data.frame*, which is appended to the return object. This allows further investigation, sorting, and filtering to find the best formula with one, two, three, or any desired number of variables. In the end, like with the original function, the formula with the best fit i.e., the one with the lowest waic is returned. This advantage is of course gained by a much longer computing time depending on the number of variables available and the number of variables allowed through *include_max* and their possible combinations. For explanations of the other input parameters I refer to the documentation of the original *INLAstep()* function within the package *INLAutils* (Redding et al. 2017).

3.4 Model selection

As presented in the previous sections, 16 functions that in sum produce 67 numerical parameters describing the river properties on a hectometer resolution, were produced and the results collected within a database. Afterwards combinations of these parameters were put into INLA models. To avoid overfitting, a maximum of three variables was included within each formula and to reduce computing time, every variable, that was available with different buffer margins, was firstly tested with a 1000 m buffer only. This left an overall number of 6017 different formula combinations that were tested individually via the fork of the *INLAstep()* function.

To further improve the resulting models three additional steps were taken. Firstly, random effects were integrated into INLA via the *f()* function and secondly, spatial effects were taken into account via the SPDE approach (see section 3.1.2). The inclusion of random effects into the model is done via the sampling ID. This allows the *inla()* function to vary the intercept and the

slope of the fitted model to a certain degree which should improve the model fit. Thirdly, the best ten models for the three tested time periods, that had parameters with adjustable buffers in them, were tested for every buffer or combination of buffers to further improve the models. This, however, yielded no improvements and is therefore not further discussed within this work.

4 Results

Presented within this chapter are the results of the erosion rates, the queried predictor variables, and finally the derived inla models. The significance of the produced models as well as comparisons between the models incorporating only the parameters, parameters and random effects, and the models containing parameters, random effects, and the SPDE approach are shown.

4.1 The erosion rates for the different time periods

The earliest and longest time period ranging from 1898 to 1995 shows two peaks around the gauging stations Torgau and Magdeburg Strombrücke (Fig 10 A). The maximum value for the erosion rate within this time period is 3.4 cm/a, reached at river-km 155. The minimum value is -3.2 cm/a i.e., an aggradation of 3.2 cm/a, reached at river-km 459. The mean erosion rate is 0.69 cm/a for the whole stretch until the barrage of Geesthacht. For the area of the erosion reach between river-km 120 and river-km 290 the mean erosion rate 1.07 cm/a. Time period 1898–1995 shows the least variation of erosion rate values of the three time periods.

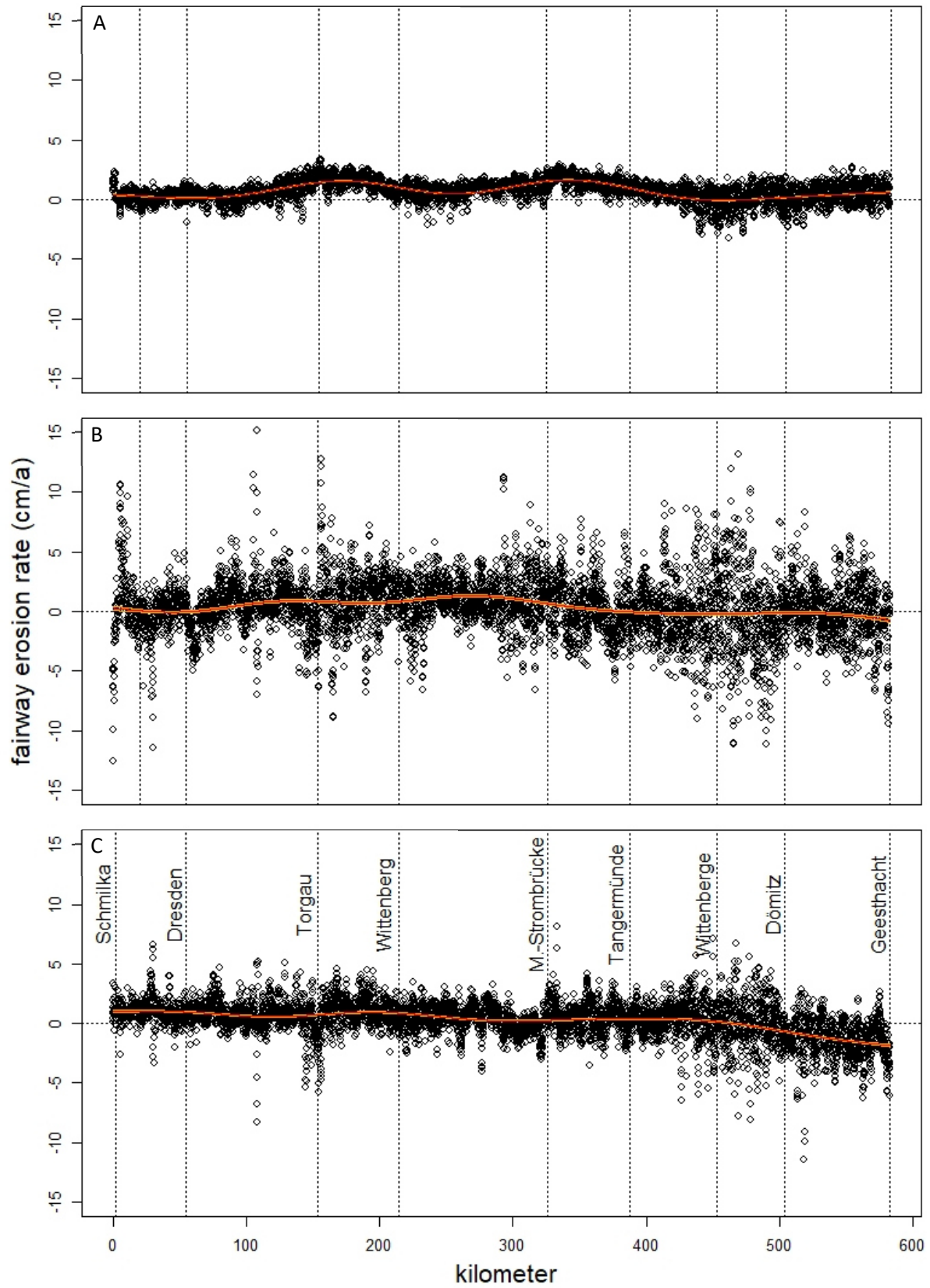


Figure 10 The annual fairway erosion rates plotted against the kilometrage of the Elbe of the time periods 1898–1995 (A), 1995–2004 (B), and 2004–2017 (C).

For time period 1995–2004 the mean erosion rate within the erosion reach is 0.93 cm/a with a maximum of 12.76 cm/a (Fig. 10 B). The overall maximum erosion rate for this time period is 15.11 cm/a at river-km 108, the minimum erosion rate is -12.59 cm/a within the first river kilometer. The mean erosion rate is 0.35 cm/a. This time period shows the highest variation in erosion rates. These variations are apparent throughout the whole course of the river but rise in strength from river-km 400 onward with a maximum difference of 23 cm/a in erosion rate between river-km 463 and 464. The last 50 km before the barrage of Geesthacht are characterised by aggradation rather than erosion.

The time period 2004–2017 is also characterised by a larger variation within erosion rates than time period 1898–1995 but shows lesser variation than time period 1995–2004 (Fig. 10 C). The maximum difference between two adjacent river kilometers is 12.33 cm/a between river-km 517 and 518. Again, the variation in erosion rates strengthens from river-km 400 onwards. The mean erosion rate within the erosion reach is 0.73 cm/a, the mean of the whole river between Schmilka and Geesthacht is 0.27 cm/a. The maximum erosion rate is 8.13 cm/a at river-km 333. The minimum erosion rate is -11.42 cm/a at river-km 518. Downstream of the gauging station Wittenberge the Elbe is characterised by aggradation which successively increases until the barrage of Geesthacht.

4.2 Queried hydrogeomorphic variables

This section presents the results of selected characteristic or modelling relevant query functions along the Elbe. A comprehensive overview over all the function results plotted along the river axis can be found in the Appendix.

4.2.1 Minimum height of the DEM

The minimum height of the DEM characterises four parts of the Elbe (Fig. 11): (I) The first part ranges from the rivers origin where it enters Germany to about river-km 160 around the gauging station Torgau. In this first part the slope of the riverbed is at its steepest and appears to be almost perfectly linear. Reaching Torgau the slope becomes abruptly flatter entering the second section of the river after passing the Torgauer Felsen. (II) This

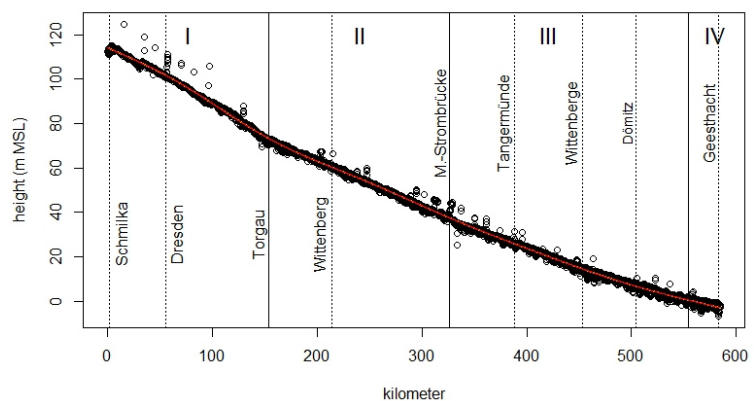


Figure 11 The minimum height of the DEM plotted against the course of the Elbe. Additionally, the four characteristic parts (I gravel bed river, II & III middle parts, IV backwater of Geesthacht) of the river are marked, divided by solid vertical lines.

second part spans from Torgau to the gauging station Magdeburg-Strombrücke. Here, a

second bend in the slope is visible just after the erosion barrier Domfelsen, again levelling out the slope of the riverbed to some degree. (III) Downstream of Magdeburg the slope remains constant up until reaching the impounded part of the river that is fully regulated by the barrage of Geesthacht. (IV) This part begins between Wittenberge and Dömitz around river-km 560 and is indicated in the graph by another bend levelling out the slope even more. At Geesthacht the reach of the Binnenelbe comes to an end. These four characteristic parts of the river are referred to throughout this work.

4.2.2 Difference in river axes and river slope

Positive returns of *getAxisADiff()* indicate that the historical axis was longer than the recent one i.e., the length of the axis shortened over time. Parts of the river where the axis elongated over time have negative values. Most of the river did not change in its length since 1830 with some spatially

distinct exceptions (Fig. 12 A). Just before the gauging station Torgau until the gauging station Wittenberg most queries indicate a shortening of the axis. The same is true for the stretch just below the gauging station Magdeburg Strombrücke and for a small part of the river between the gauging stations Tangermünde and Wittenberge. A significant elongation of the river axis

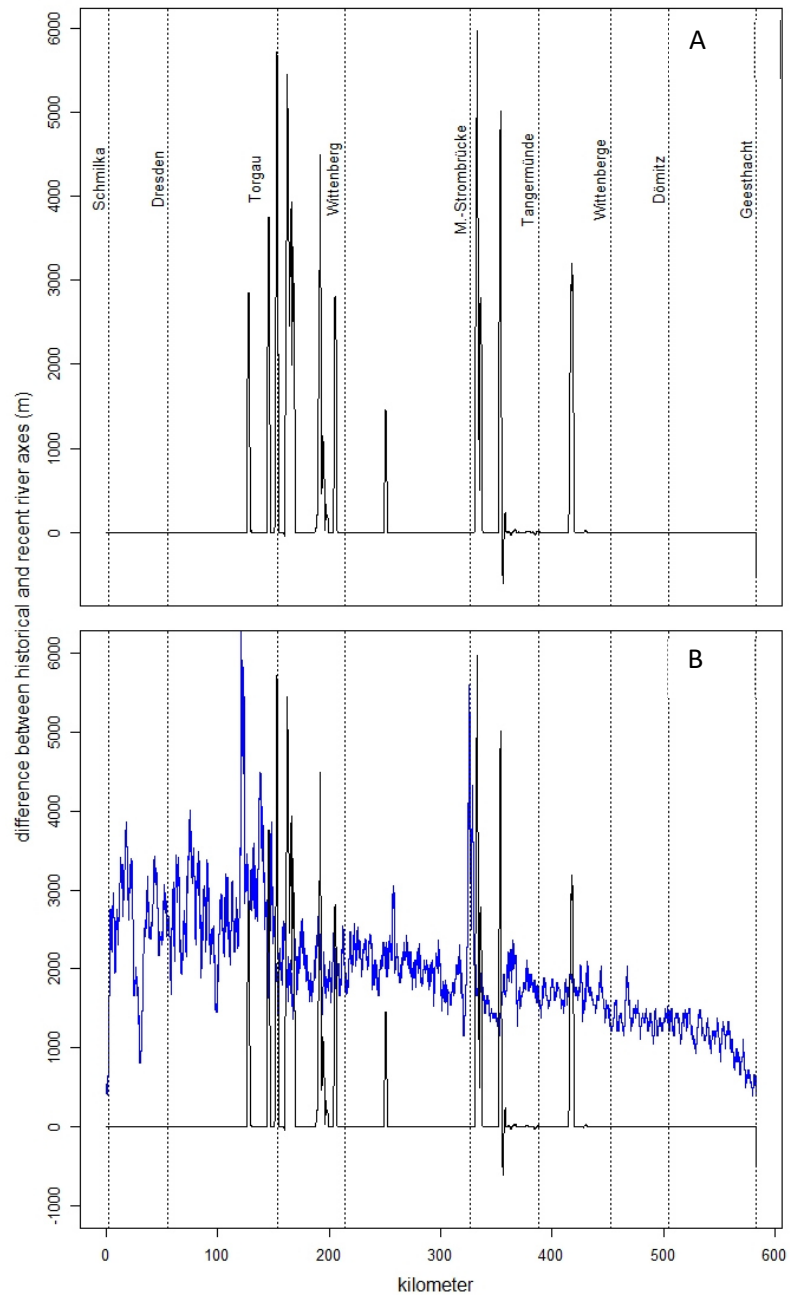


Figure 12 (A) The changes in river axis length plotted against the course of the Elbe. (B) The changes in river axis plotted along the outputs of the function *getSlope()*. The peaks of the slope occur just before the peaks of the changes of the axis. Here, the slope is depicted inversed i.e., larger values indicate a steeper slope and multiplied by 10 Mio. to illustrate the effects more clearly.

can only be seen in a short stretch between river-km 355 and 385 and just before the barrage of Geesthacht from river-km 582 onward.

The parts where a shortening of the axis occurred over time are met with an increase in the slope (Fig. 12 B). Within the slope curve two peaks can be seen that both occur just before the parts of the river that are characterised by the greatest axis shortening. The peaks of the erosion lie at river-km 121 and 325 which is 34 km before the erosion barrier Torgauer Felsen and directly where the erosion barrier Domfelsen is located (see section 2.2).

4.2.4 Proportions of active floodplain between MLQ and MHQ

Within the first part of the river only about 20 % of active floodplain get additionally covered by water during times of high discharge (Fig. 13 A). In the second and third part of the river, as defined in Fig 12, 70–90 % get additionally covered. In part IV this value drops rapidly to zero.

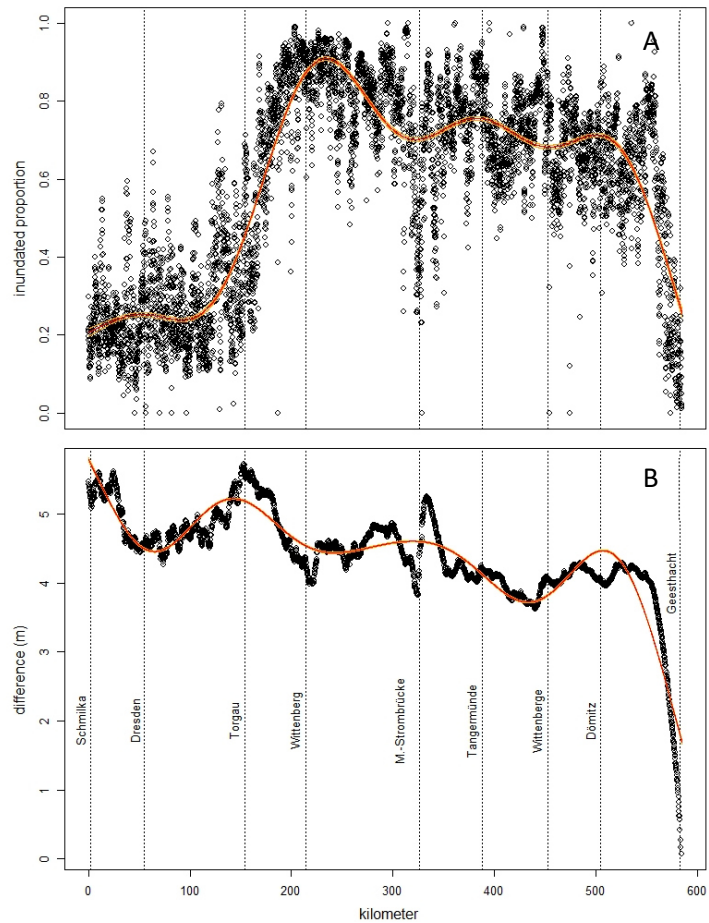


Figure 13 The proportion of the active floodplain that is additionally inundated when the discharge scenario changes from MLQ to MHQ (A). The difference in water level between MLQ and MHQ plotted against the kilometrage of the river (B).

4.2.5 Height differences between FLYS3 water levels MLQ and MHQ

The difference in water level between the two discharge scenarios MLQ and MHQ is at its greatest within the first part of the river and then gets progressively smaller (Fig. 13 B). In the regulated last part, the water levels do not change at all. Two peaks at gauging station Torgau and just below the gauging station Magdeburg Strombrücke can be seen with a difference of 5.73 and 5.25 m respectively. The mean difference is 4.41 m.

4.2.6 Proportion of PNV type hardwood floodplain wet (HFW)

The proportion of PNV type hardwood floodplain wet (HFW) is 1.46 % on average (Fig. 14). This type of vegetation is hardly detected until river-km 350. Here, the proportion increases and reaches its maximum at river-km 552 with 34.5 %.

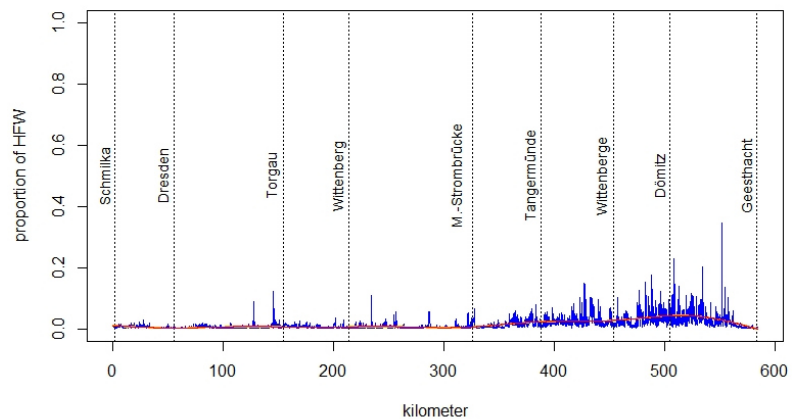


Figure 14 Proportions of the active floodplain of type HFW the course of the Elbe i.e., the proportion inundated between 49 and 79 days per year..

4.2.7 Proportion of PNV type reed (RE)

PNV type reed (RE) starts occurring frequently just before the gauging station Wittenberg and from then on continues to occur along the whole river (Fig. 15). The proportion covered is large compared to the vegetation type HFW. On average 5.91 % of the active floodplain are characterised by RE. The maximum is reached at river-km 481 with 39.1 % of the respective cross-section area.

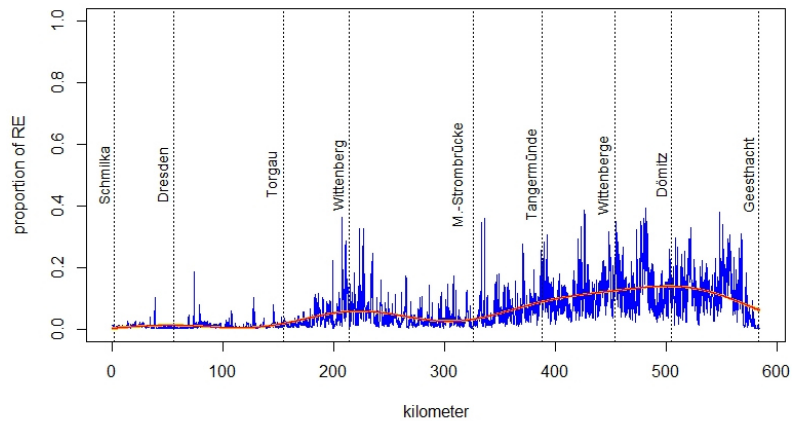


Figure 15 Proportions of the active floodplain of type RE along the course of the Elbe.

4.2.8 Proportion of PNV type periodically flooded (PF)

The first part of the river is the area with the smallest proportions of PNV type periodically flooded (PF) (Fig. 16). Further downstream PF becomes the main type of vegetation for large parts of the active floodplain. The area covered by PF is 49.7 % with some cross-section areas covered to over 90 %. The maximum is reached at river-km 238 with 93.7 % PF.

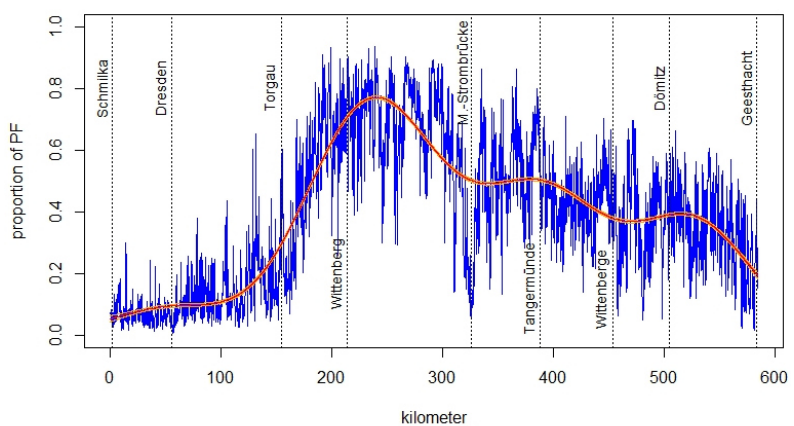


Figure 16 Proportions of the active floodplain of type PF along the course of the Elbe.

4.2.9 Proportion of DLM25 landuse type grassland

The DLM25 landuse type grassland is present along the whole course of the river and takes up large proportions of the active floodplain (Fig. 17). On average, this type of landuse takes up 39.18 % of the active floodplain with some areas being almost completely covered.

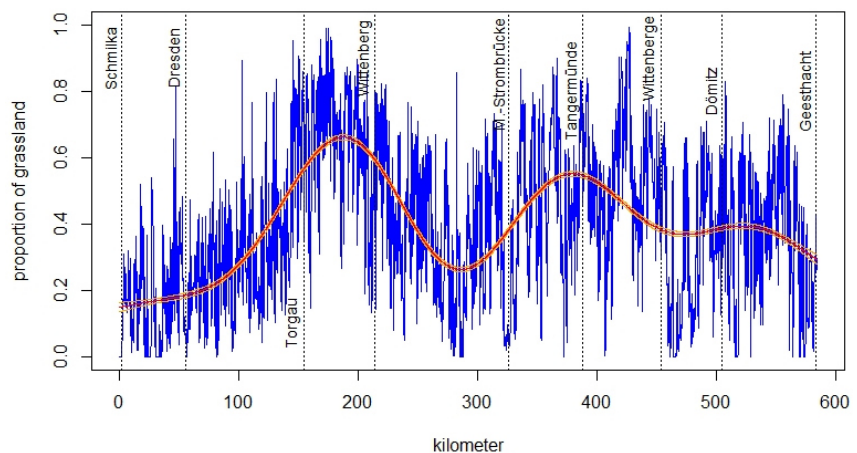


Figure 17 Proportions of the active floodplain containing the DLM25 type grassland along the course of the Elbe.

The maximum is reached with 99.26 % at river-km 427. Although grassland is apparent along the whole course of the river, two peaks at river-km 180 and river-km 370 can be seen.

4.2.10 Proportion of DLM25 landuse type human

Human altered landuse can be found to a great proportion in the first part of the river (Fig. 18). Here, in a narrow valley with steep mountains, large areas of the active floodplain are covered by human infrastructure to more than 80 % with a maximum of 91.01 % at river-km 35. Apart from this, no real trend in the proportion of human altered parts of the active floodplain can be seen along the whole course of the river. On average, this type of landuse takes up 21.32 % of the active floodplain.

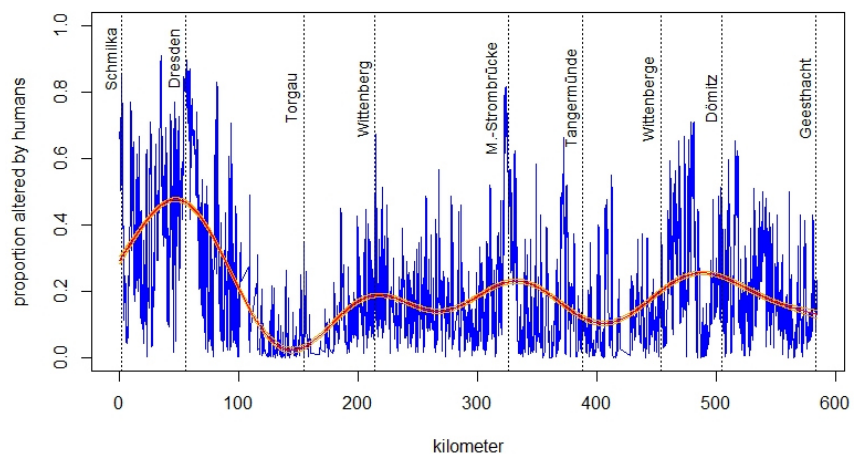


Figure 18 Proportions of the active floodplain containing the DLM25 type human along the course of the Elbe.

4.3 Best models

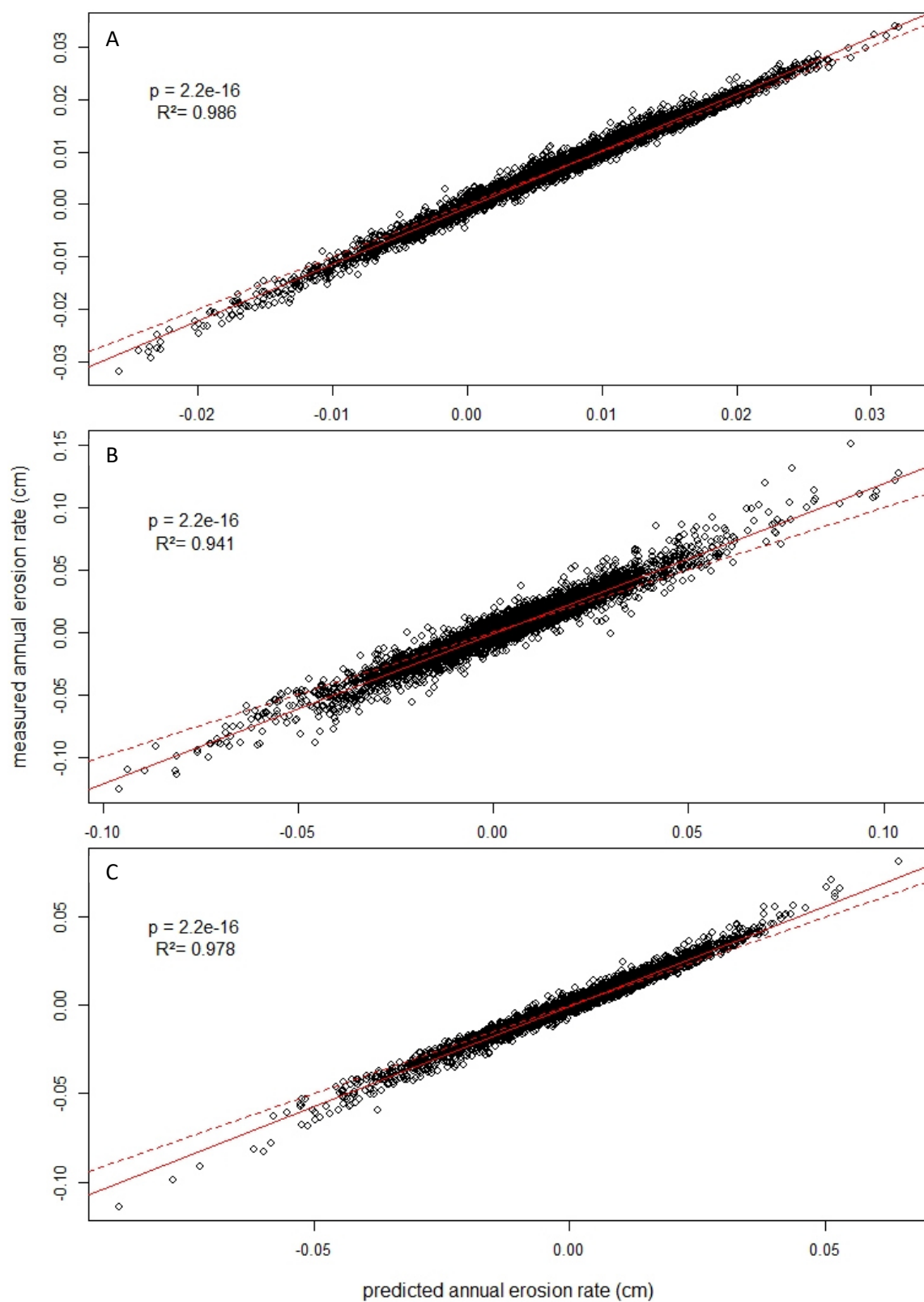


Figure 19 Erosion rates for the three time periods (1898–1995 (A), 1995–2004 (B), 2004–2017 (C)) plotted against the fitted erosion rates for the respective best models. For each time period a regression line was calculated via the function `lm()` and is shown as red line. The dashed line within the plots depicts the theoretical perfect fit with slope 1. Also, the p - and R^2 -values are presented.

4.3.1 1898–1995

The searches for the best formula for the first time period between 1898 and 1995 yielded equation (1).

$$m1 < -DA_{1898_1995} \sim Intercept + DiffbetweenMNQMHQ + pVegType_{HFW} + pVegType_{RE} + f(id, model = 'iid') + f(spatial.field, model = spde) (1).$$

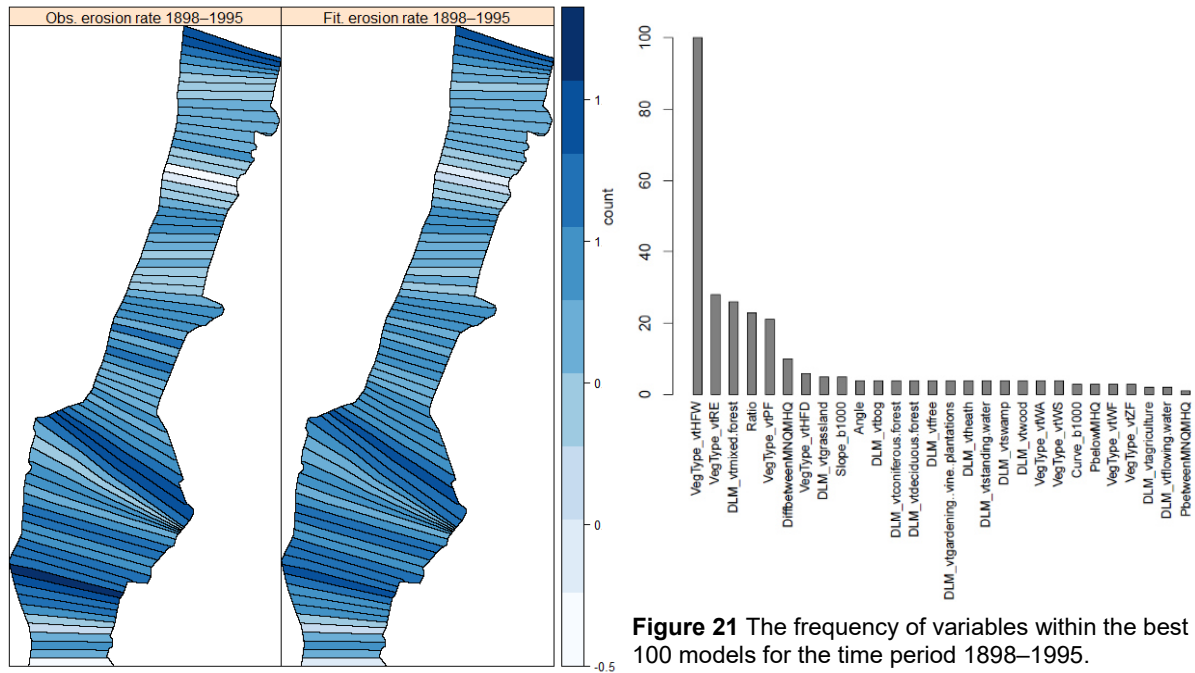


Figure 21 The frequency of variables within the best 100 models for the time period 1898–1995.

Figure 20 Depiction of the cross-section areas of *spdf.afe_csa* for the tile *e033_TANGERMUENDE* coloured based on the erosion rate (cm/a) for the time period 1898–1995. On the left-hand side the measured erosion rates are depicted and on the right-hand side the erosion rates fitted by model *m1* are shown.

The parameters that fit the annual erosion rate within the time period 1898–1995 best are the difference between the water levels MLQ and MHQ (*DiffbetweenMLQMHQ*), the proportion of PNV type hardwood floodplain wet (*pVegType_HFW*), and the proportion of PNV type reed (*pVegType_RE*). The fitted values match the measured values of the annual erosion rate to a great degree (Figs. 19 A & 20).

The low p-value of $2.2e-16$ for the linear regression model between fitted and measured values shows that their dependency is highly significant (see also Fig. 23). The linear regression model lies below the ideal curve for areas with negative erosion rates and above the ideal curve for positive erosion rate values, indicating overestimation for positive values and underestimation for negative values.

A frequency analysis of predictors in the best 100 of the 6017 possible inla models reveals that there are five variables that seem to have the highest explanatory power because they appear in more than 20 models each (Fig. 21). These are PNV types periodically flooded (*VegType_PF*), reed (*VegType_RE*), and hardwood floodplain wet (*VegType_HFW*), DLM 25

type mixed forest (*DLM_mixed.forest*), and Ratio. *VegType_HFW* occurs in every one of the best 100 models

When evaluating the waic of the three different model types (parameters only, parameters and random effect, parameters, with random effect and SPDE), three distinct distributions can be seen (Fig. 22). The models only including the parameters are the worst i.e., have the highest waic, followed by better models including random effects, and then by again better models including both random effects and the SPDE approach. The models from the different approaches are distributed very narrowly and the difference in result among the approaches is comparatively large.

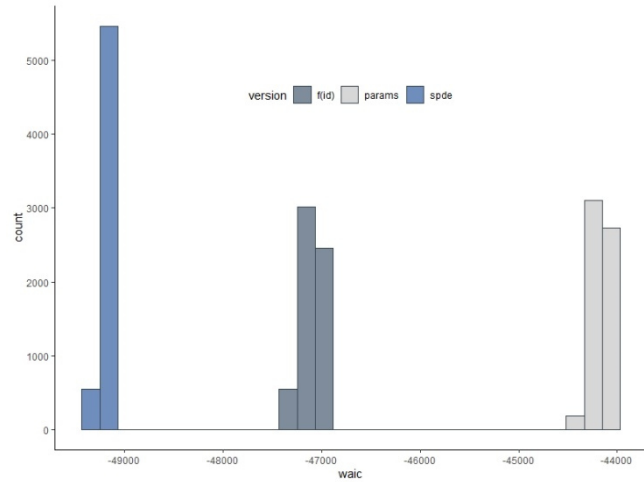


Figure 22 Comparison of the waic of the models containing only the parameters (params), parameters and random effects (f(id)), and parameters, random effects and SPDE (spde) for the time period 1898–1995.

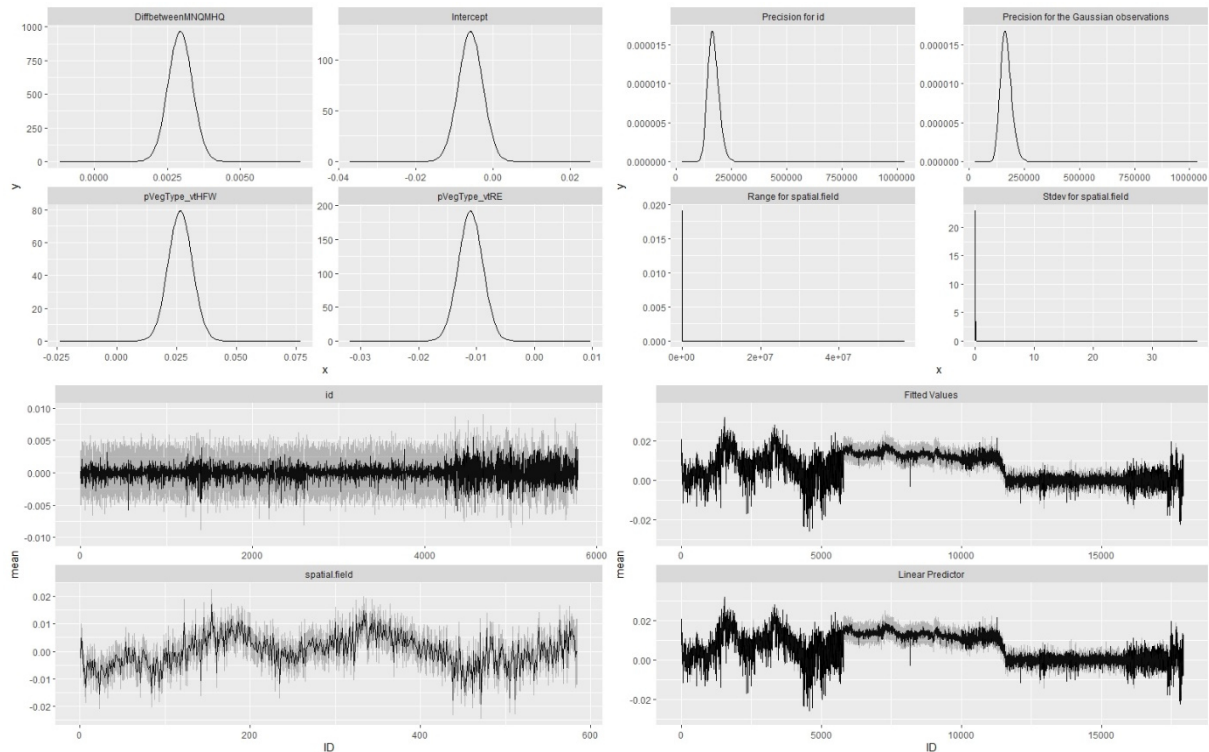


Figure 23: Graphical summary of model m1. Four plots top-left: marginal posterior distributions, four plots top-right: marginal posterior distributions, two plots bottom-left: marginal posterior mean and 95% credible intervals of the spatial and random effects, two plots bottom-right: marginal posterior mean and 95% credible intervals of the linear predictor and fitted values.

4.3.2 1995–2004

The searches for the best formula for the time period 1995–2004 yielded equation (2).

$$m2 < -DA_{1995_2004} \sim Intercept + Slope_{b1000} + pVegType_{PF} + pDLM_{agriculture} + f(id, model = 'iid') + f(spatial.field, model = spde) \quad (2).$$

The parameters that show the best fit of the annual erosion rate within the time period 1995–2004 are the average slope over a longitudinal buffer of 1000 m ($Slope_b1000$), the proportion of the PNV type periodically flooded ($pVegType_PF$), and the proportion of the DLM25 type agriculture ($pDLM_agriculture$). The fitted values match the measured values of the annual erosion rate to a great degree (Figs. 19 B & 25).

The low p-value of $2.2e-16$ for the linear regression model between fitted and measured values shows that their dependency is highly significant (see also Fig. 24). The linear regression model lies below the curve for the ideal fit for negative erosion rates and above it for positive erosion rates, again indicating an overestimation for positive values and an underestimation for negative values.

The frequency analysis of predictors in the best 100 models showed none of the variables is dominant here (Fig. 26). There are ten variables that appear in more than ten and only one that appears in more than a quarter of the best formulas. When looking at the best ten models

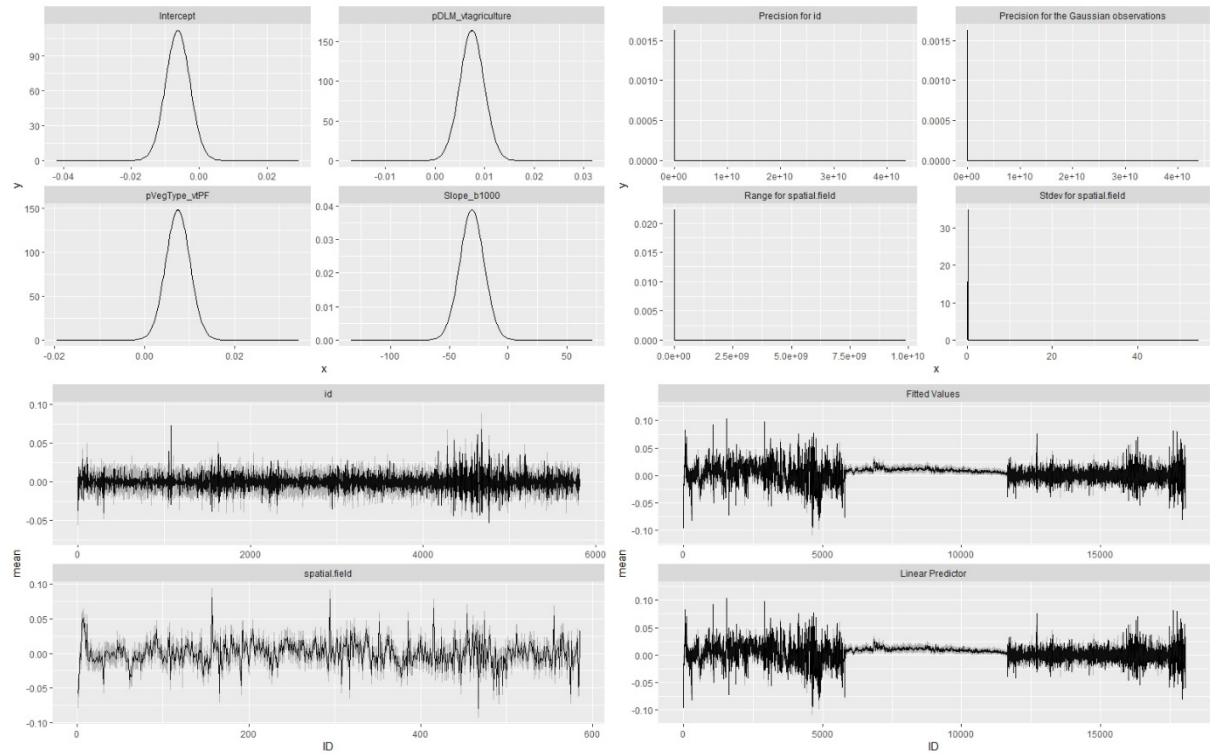


Figure 24 Graphical summary of model m2. Four plots top-left: marginal posterior distributions, four plots top-right: marginal posterior distributions, two plots bottom-left: marginal posterior mean and 95% credible intervals of the spatial and random effects, two plots bottom-right: marginal posterior mean and 95% credible intervals of the linear predictor and fitted values.

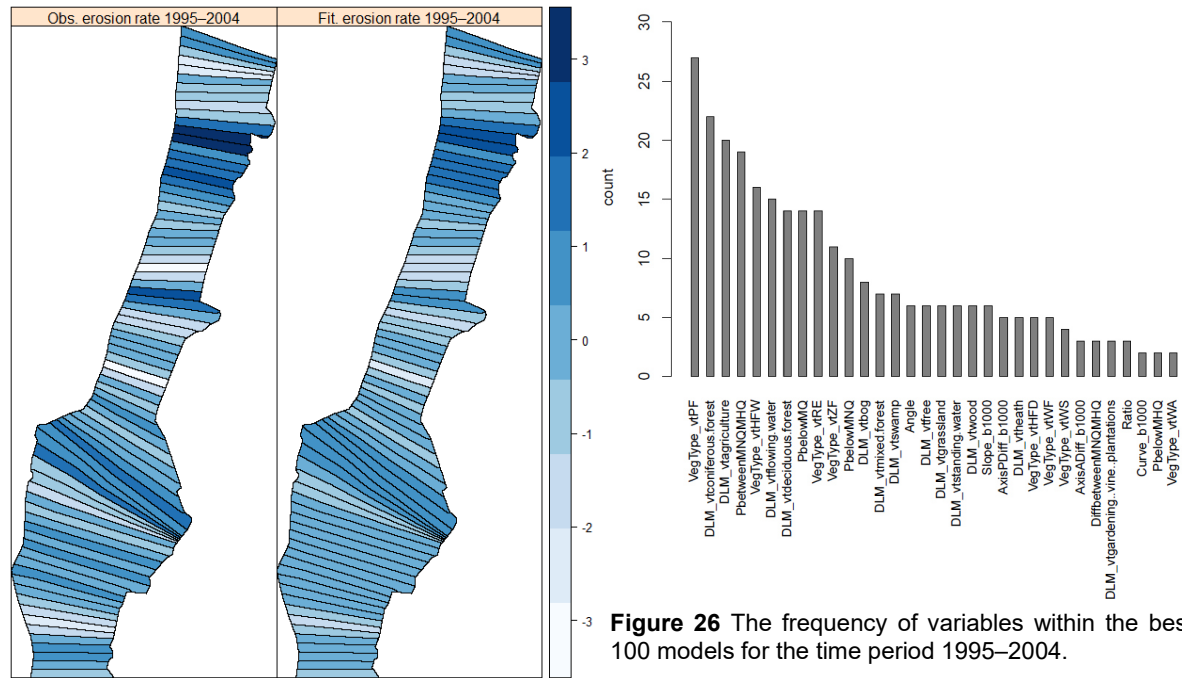


Figure 25 Depiction of the cross-section areas of *spdf.afe_csa* for the tile *e033_TANGERMUENDE* coloured based on the erosion rate (cm/a) for the time period 1995–2004. On the left-hand side the measured erosion rates are depicted and on the right-hand side the erosion rates fitted by model m2 are shown.

however the two variables PNV type periodically flooded (*VegType_tvPF*) and DLM25 type coniferous forest (*DLM_coniferous forest*) stand out more clearly, having six appearances each. Despite being within the best and second-best formula, the variable describing the average slope over a 1000 m buffer (*Slope_b1000*) only appears a total of six times within the best 100 formulas.

The difference between the models including random effects and models only including the parameters is comparatively small with a difference of about 1500 in waic whereas the inclusion of the SPDE approach seems to have a great effect, lowering the waic by about an additional 3000 points as compared to the random effect models (Fig. 27).

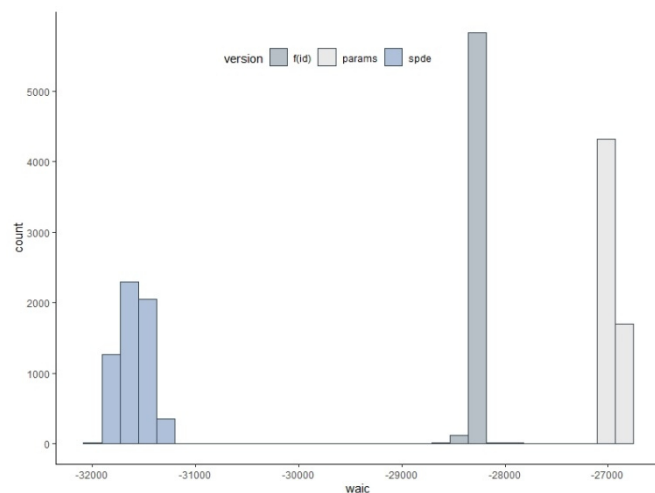


Figure 27 Comparison of the waic for the models containing only the parameters (params), parameters and random effects (*f(id)*), and parameters, random effects and SPDE (*spde*) for the time period 1995–2004.

4.3.3 2004–2017

The searches for the best formula for the time period 2004–2017 yielded equation (3).

$$m3 < -DA_{2004_{2017}} \sim Intercept + Slope_{b1000} + pDLM_{grassland} + pDLM_{human} + f(id, model = 'iid') + f(spatial.field, model = spde) \quad (3).$$

The parameters that fit the annual erosion rate of this time period are the average slope over a 1000 m longitudinal buffer (*Slope_b1000*), the proportion of the active floodplain covered by DLM25 type grassland (*pDLM_grassland*), and the proportion covered by human infrastructure (*pDLM_human*). The fitted values match the measured values of the annual erosion rate to a great degree (Figs. 19 C & 28).

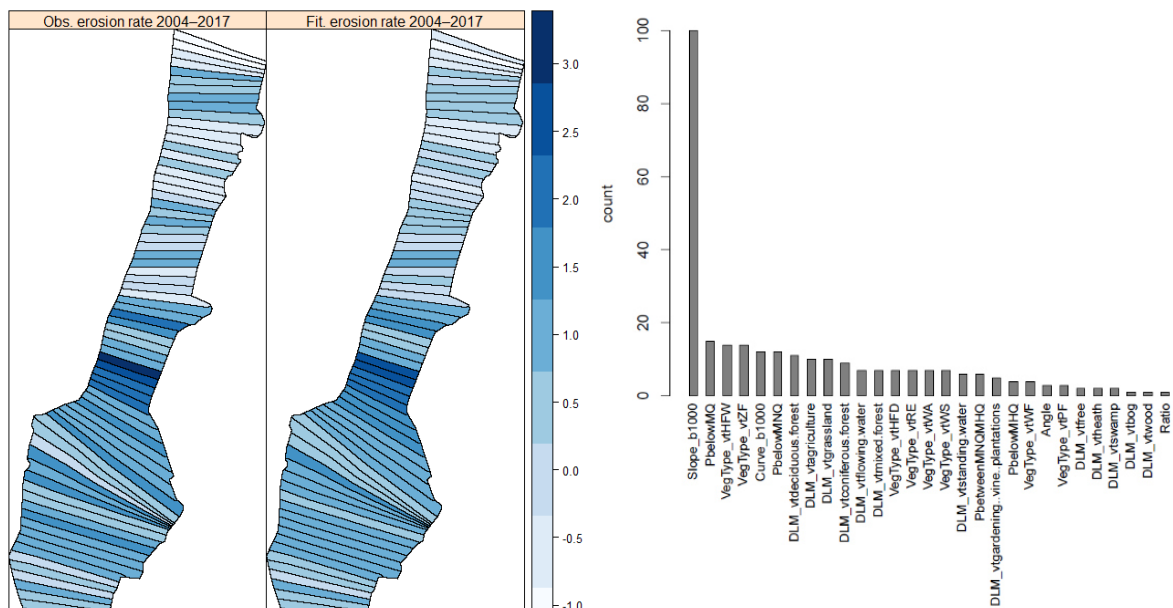


Figure 29 The frequency of variables within the best 100 models for the time period 2004–2017.

Figure 28 Depiction of the cross-section areas of *spdf.afe_csa* for the tile *e033_TANGERMUENDE* coloured based on the erosion rate (cm/a) for the time period 2004–2017. On the left-hand side the measured erosion rates are depicted and on the right-hand side the erosion rates fitted by model m3 are shown

The low p-value of 2.2e-16 for the linear regression model between fitted and measured values shows that their dependency is highly significant. The linear regression model lies again below the ideal curve for areas with negative erosion rates and above the ideal curve for areas for positive erosion rates (see also Fig. 31).

The frequency analysis of the clearly shows that *Slope_b1000* is the best predictor for the erosion rate within this time period (Fig. 29). It is present within every formula while none of the other variables appear in more than 20 %.

This time period shows the smallest margins between the three distributions of the three different modelling methods (Fig. 30). Here, the distributions almost overlap with only about 1000 waic points between the parameter and random effect models as well as the random effect and the SPDE models. The distribution of the SPDE models, however, is very broad, ranging from about -36000 to about -40000 in waic with few models showing very high fit.

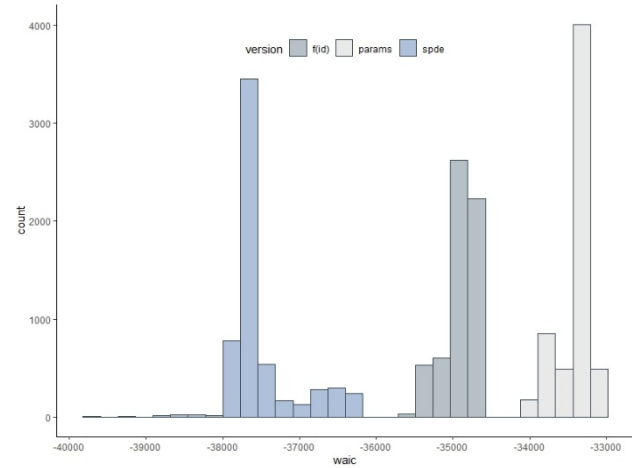


Figure 30 Comparison of the waic for the models containing only the parameters (params), parameters and random effects (f(id)), and parameters, random effects and SPDE (spde) for the time period 2004–2017.

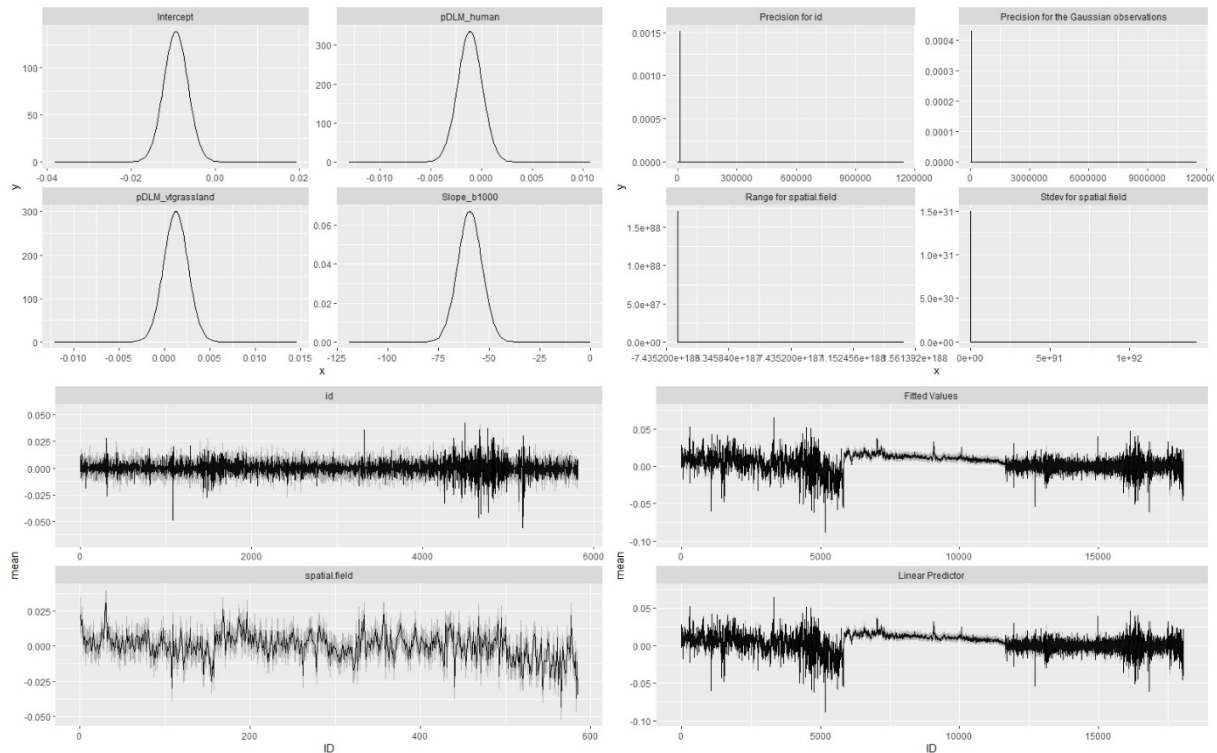


Figure 31 Graphical summary of model m3. Four plots top-left: marginal posterior distributions, four plots top-right: marginal posterior distributions, two plots bottom-left: marginal posterior mean and 95% credible intervals of the spatial and random effects, two plots bottom-right: marginal posterior mean and 95% credible intervals of the linear predictor and fitted values.

5 Discussion

The INLA approach yielded regression models with satisfactory results for every analysed time period. All the fitted erosion rates match the respective measured erosion rates to a great degree, even though they differ in their individually selected predictor variables. These

differences, however, can probably be related to time specific differences in data quality and sediment management carried out by the waterway and navigation authorities.

For all the time periods, the fitted erosion rate values slightly overshoot the measured values i.e., negative erosion rates are predicted smaller than the measured values and positive erosion rates are predicted larger than they were measured. Nevertheless, the presented approach is considered to work i.e., fits the measured erosion rates to a satisfactory high degree and therefore proves the applicability of INLA in environmental modelling contexts.

Apart from INLA, there are other, more established but computationally much more expensive, ways of accounting for autocorrelation within spatial data. The most common way of doing this, is using MCMC (Markov chain Monte Carlo) methods (Gómez-Rubio 2020). The main argument towards using MCMC is that it is asymptotically exact, whereas INLA is a bit more approximate (Taylor & Diggle 2012). Due to the fact however, that the amount of data and the spatial region covered within this work were very large, it was supposed that the fitted values calculated by the INLA method are also satisfactorily accurate and the use of MCMC would exceed the available computational power by far. Therefore, the INLA method was used.

What this work lacks and what can be added during future research are two things. The first is the inclusion of sediment properties to capture the heterogeneity of the riverbed. As mentioned within section 2.2 the composition of the riverbed influences the erosion patterns within a river to a great degree. Unfortunately, by the time this study was taken, there were no data riverbed available in a resolution that would have suited the other data used within this work. However, as soon as the data describing the riverbed composition is available, it can easily be added to the analyses and is bound to further improve our understandings of the drivers behind a river's erosion. The second thing, that can be used to further improve the results in the near future, is the integration of a temporal component into the analyses. The erosion rate of the Elbe nowadays gets measured periodically via echo sounding. The results of this monitoring could be used to add a time dimension to the analyses and is also bound to further improve the study results.

One last thing, that I would like to stress here, is that the models were created only by fitting existing data from different environmental parameters to existing data of the erosion rates in the best possible way. Therefore, the results were strictly speaking not *predicted*. Another, more substantive approach to modelling the erosion rates would have been to only use a part of the measured data to fit the model and then try to use the model to predict the rest of the measured data. This would have been a more rigorous way to verify the predictive capacities of the model. This, of course, is also realisable during future research.

5.1 Results of the query functions

The findings of the query function *getPbetweenMLQMHQ()* can be explained by the geology of and around the river. In the first part the river Elbe flows through the Elbe Sandstone Mountains. Here, the river is flanked by high and steep rock walls and therefore has not much space to form a wider floodplain. This means that in times of high discharge almost all the additional water is converted to a greater depth of the river rather than to a greater width resulting in an only slightly bigger area covered by water. Downstream from Dresden the valley gets progressively wider and thus the difference in area coverable by water during high discharge gets bigger. This greater difference is observable until the last part of the Elbe, where the river is completely regulated by the barrage of Geesthacht. Here, the discharge is regulated and therefore the area covered by water remains constant. These explanations can also be transferred to the findings of *getDiffbetweenMLQMHQ()*. Within the steep valleys of the first river part, small amounts of additional water cause a relatively large difference in water level, whereas even a great amount of water does not change the water level much, when the water has great areas flanking the river into which it can flow. The last part around the barrage of Geesthacht is regulated. Therefore, there is hardly ever a change in water level.

The apparent commonalities between areas with a large deficit between the historical and recent river axis and areas characterised by steep slopes, that can be seen in Fig. 12 B are plausible. Wherever a large amount of axis shortening took place the river has less horizontal distance to cover the same vertical distance and therefore has a greater slope. It also makes sense that a steeper slope, represented by the variable *Slope_b1000*, is associated with higher erosion rates since the river flows with a higher velocity at this point. This higher velocity comes both from the shortened path that must transport the same amount of water per time and from the steeper slope that naturally leads to higher flow velocities. It is therefore generally understandable, why the variable slope plays such an important role in explaining the erosion rates in so many of the best formulas, especially for the most recent time period. With a steeper slope and the associated higher velocity, the river has a higher competence which ultimately leads to a higher rate of erosion (Gilbert 1914).

The competence of a river also increases, when the discharge Q increases, which is depicted by the returns of the function *getDiffbetweenMLQMHQ()* (Fig. 13 B). Q is defined as the amount of water per cross-section area of the river per time. Therefore, it does not increase as much when the river has wide space to stretch out in times of rising water levels (Gilbert 1914). However, when the river does not have this space i.e., there is a large difference between the water levels MLQ and MHQ , Q increases.

5.2 Time period 1898–1995

The time period 1898–1995 has to be discussed separately from the other two time periods. This is due to different reasons. For one thing, the time period spans much longer than the other two, integrating data over 97 years whereas the other two time periods only gather data from nine and 13 years respectively. Therefore, short-term changes in erosion patterns average each other out which leads to lesser variability in erosion values and to smaller erosion or aggradation rates in general (Fig. 10). The maximum erosion rate lies at 3.4 cm/a and the overall erosion rate averages at 0.69 cm/a. This means that the riverbed from 1995 onward already lies on average almost 70 cm below the riverbed present at the start of measurements in 1898 and some parts of the riverbed may even have dropped over 3 m. This change in riverbed height may have altered the properties of the riverbed and therefore makes the datasets hardly comparable. Another aspect, why this early time period cannot be compared reasonably to the other two time periods, is the difference in data acquisition and data quality. For the time period 1898–1995 the data quality is assumed to be comparatively low, considering that the data were not collected by echo sounding but by consulting the literature i.e., the “Elbstromwerk” (see section 3.2.9).

The three best parameters for this time period are the difference in water level between MLQ and MHQ (*DiffbetweenMLQMHQ*), the proportions of the active floodplain that are of PNV type hardwood floodplain wet (*pVegType_HFW*), or PNV type reed (*pVegType_RE*). As discussed in section 5.1 it seems reasonable that a higher *DiffbetweenMLQMHQ* is associated with higher erosion rates because it is ultimately a measure of how much the river’s surroundings resemble a canyon. This canyon-like surrounding can either be the cause or the result of higher erosion rates. Thus, it is thought to be an environmental parameter always correlated to high erosion rates regardless of whether it drives this process or results from it. Given the large span of the time period, it was inevitable for this parameter to show a strong correlation.

The other two parameters influencing the erosion rates to a great degree are the proportions of the active floodplain of the PNV types hardwood floodplain wet (*pHFW*) and reed (*pRE*). These parameters both resemble the proportion of the active floodplain inundated for a certain amount of time per year. In the case of *pHFW* the floodplain is inundated between 50 and 80 days per year. In the case of *pRE* the inundation frequency lies between 159 and 219 days per year. If the proportion of *RE* is high, this means that a large part of the active floodplain is inundated for most of the time. This in turn indicates the same as a large value for *Pbetween-MLQMHQ*, namely that there is little floodplain left where additional water can flow off to. Therefore, the water runs in the same course most of the year leading to a high rate of erosion. The proportion of the floodplain that is of type *HFW* resembles the area that gets inundated a few

weeks per year, when the water rises above the canyon-like structure indicated by high values of $P_{betweenMLQMHQ}$ and pRE . When the river has larger areas to flow in, this usually slows down the water column and leads to aggradation rather than to erosion. The pattern shown here, indicating higher erosion rates when $pHFW$ is large, may be due to the larger amount of water flowing in the RE areas and causing erosion rates that are larger in absolute value than the aggradation rates within the HFW areas.

5.3 Time periods 1995–2004 and 2004–2017

The data collection of the time periods 1995–2004 and 2004–2017 took place from 1993 to 1999 for the year 1995, from 2003 to 2004 for the year 2004, and from 2016 to 2017 for the year 2017 via echo sounding, representing the first official data after the German reunification. Therefore, the data is much more reliable than the data discussed in the previous section. Also, the time spans, the data were collected in, were much smaller which allows for unusual events like exceptional floods to manifest themselves.

The best predictors for the erosion rates within time period 1995–2004 are the average slope over a longitudinal buffer of 1000 m ($Slope_b1000$) and the proportions of the active floodplain that are covered by the PN type periodically flooded ($pVegType_PF$) and the DLM25 type agriculture ($pDLM_agriculture$). The association of the proportion characterised by agriculture is probably explained by the fact that the river is characterised by high erosion rates within this region. Therefore, the riverbanks are steep, and the inundation probability of crop is little while agriculture still has the advantage of using the rich soils within the floodplain. Therefore, agriculture is not the driver of the high erosion rates but a result from them. Regardless of if the parameter is responsible for the erosion or not, it still represents a good proxy for areas with high erosion rates. $pVegType_PF$ (pPF) quantifies the proportion of the active floodplain inundated between 219 and 199 days per year. For this type of “vegetation” the reasoning for pRE outlined in the previous section applies even stronger. Wherever the proportion of floodplain that is almost inundated the whole year is large, so is the erosion rate. The river has hardly any space to widen in times of high discharge and therefore all additional force is translated into erosion. The last parameter with a great influence on the erosion rate within this time period is $Slope_b1000$. The influence of the slope on the erosion rate was discussed in section 5.1. In contrast to the other two parameters predicting the erosion rate for this time period, that can both also be seen as resulting from high erosion rates rather than as a cause of it, the slope of the river is a clear driving force behind erosion. It is therefore assumed to be the main reason behind high erosion rates within the years 1995 and 2004.

In general, 1995–2004 is by far the time period with the highest erosion rates and also the one with highest variations. They values reach 15.11 cm/a for the strongest erosion and 12.76 cm/a

for the strongest aggradation. These large values may be due to the extreme flood that took place in 2003. Also, this is the shortest of the three analysed time periods. Therefore, the highest variability to occur within this time period, is expectable from a statistical point of view.

The best predictor variables for the time period 2004–2017 are *Slope_b1000*, *pDLM_grassland*, and *pDLM_human*. The two DLM25 parameters can again be seen as a result of the erosion rates and probably arose from the accompanying properties of the floodplain within these areas rather than causing the erosion themselves. *Slope_b1000* is present within every of the 100 best models. It is therefore assumed, like with the time period 1995–2004, that regions of steep slopes are the main drivers for the erosion.

In general, this time period is characterised by less maximum and minimum erosion rates reaching “only” 8.13 cm/a and -11.42 cm/a respectively. 2004–2017 also shows an overall lower variability among the erosion rates. This lower variability is probably due to the longer data sampling period. The lower peaks of the erosion maxima and the overall lower erosion rates might be due to river management measures that took place within this time e.g., the addition of sediment to the erosion stretch. This may also be the cause for the strong aggradation increasing successively downstream of river-km 400.

Both time period 1995–2004 and 2004–2017 show exceptionally high variation downstream of river-km 400. This matches the description of the sediment properties by Haunschild (1996) presented in section 2.2. This finding backs the statement from section 5.1 that future analyses of erosion rates ought to integrate a high-resolution dataset describing riverbed sediment properties.

6 Conclusion

Spatial autocorrelation plays a huge role when working with environmental data or data and phenomena in general that have a spatial component. The models considering spatial relationships between the datum points were best in every aspect throughout the study. But most importantly, they yielded different results i.e., the best predictors for the erosion rates derived from the models not accounting for autocorrelation differed from them derived from the best models that accounted for it. Therefore, studies working with spatial data that do not account for spatial autocorrelation are bound to draw false conclusions.

The parameters related to landuse are not really the cause of high erosion. The different types of landuse rather arose there, where conditions of the floodplain suites them. Nevertheless, they can function as proxies to indicate such areas prone to certain degrees of erosion. The parameters with a real explanatory power were essentially the same through all three time periods. It is always areas with a high slope due to axis shortening or areas with canyon-like

structure, where the water has not enough space to widen, that show the highest erosion rates. Therefore, the most important parameters derived from this work are the difference in proportion of the active floodplain that is additionally inundated when the discharge scenario changes from MLQ to MHQ (*PbetweenMLQMHQ*), the difference in water level between the discharge scenarios MLQ and MHQ (*DiffbetweenMLQMHQ*), the absolute different in axis length between the historical and the recent river axis (*AxisADiff*), and the average slope along a 1000 m longitudinal buffer (*Slope_b1000*).

These findings suggest a clear course of action for future environmental measures aiming at a reduction of erosion rates and also mark some possible future actions of river engineering as catastrophic when aiming at conserving present-day erosion rates. The results clearly show that any means of river straightening will lead to higher degrees of erosion. Also, any means that limit the river in its ability to grow wider in times of high discharge e.g., dikes, spur dikes, and or measures of fairway lowering to improve shipping capabilities e.g., groynes, will result in higher erosion rates.

On the other hand, measures to decrease erosion rates also seem fairly obvious. In contrast to river straightening, measures that elongate the river axis would lead to lower velocities of the water column, lower competence of the river and in turn to lower erosion rates. Other measures to decrease erosion rates are the breaking of dikes, the general renaturation of the morphological floodplain or the reconnection of abandoned channels.

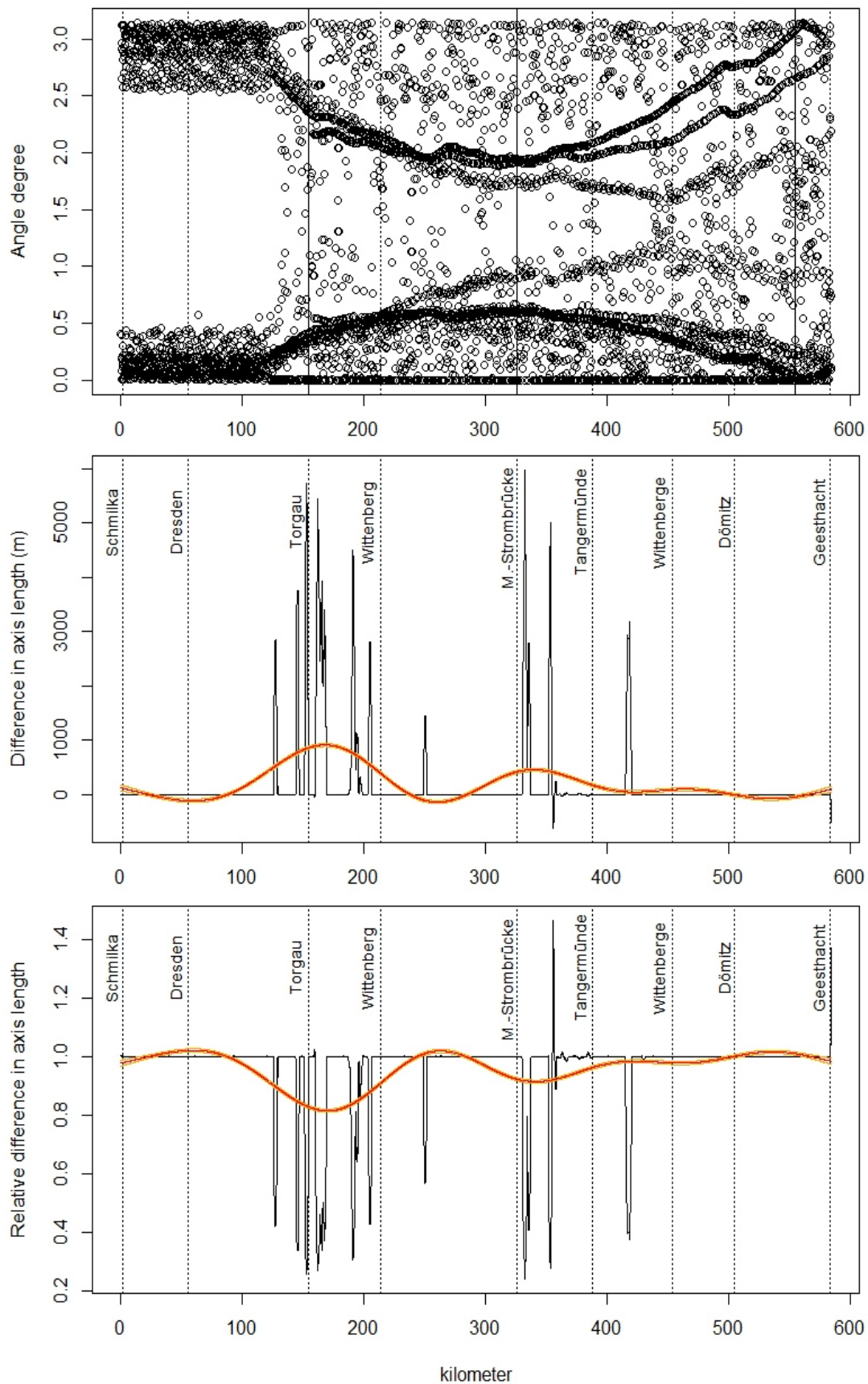
Literature

- Albery, G. (2021): Intro to modelling using INLA. Dealing with spatial autocorrelation in statistical models. Accessed 01 August 2021, <https://ourcodingclub.github.io/tutorials/inla/>.
- Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D. & F. Lindgren (2018): Spatial modeling with R-INLA: A review. In: Computational statistics, 10(6).
- Bivand, R. (2021): rgrass7: Interface Between GRASS 7 Geographical Information System and R. R package version 0.2-5. <https://CRAN.R-project.org/package=rgrass7>
- BMU [Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit] & BfN [Bundesamt für Naturschutz] (2009): Auenzustandsbericht. Flussauen in Deutschland, accessed 01 August 2021, <https://www.bfn.de/fileadmin/BfN/wasser/Dokumente/Auenzustandsbericht.pdf>.
- Bornette, G., Amoros, C., Piegay, H., Tachet, J. & T. Hein. (1998): Ecological complexity of wetlands within a river landscape. In: Biological Conservation 1998, 85,1: 35–45.
- Chun, Y., Kim, Y. & H. Campbell (2012): Using Bayesian Methods to Control for Spatial Autocorrelation in Environmental Justice Research: An Illustration Using Toxics Release Inventory Data for a Sunbelt County. In: Journal of urban affairs 2012, 34,4: 419–439.
- Elbstrombauverwaltung (Eds.) (1898): Der Elbstrom, sein Stromgebiet und seine wichtigsten Nebenflüsse (so-called Elbstromwerk). Königliche Elbstrombauverwaltung zu Magdeburg, Reimer, Berlin: 3 Vol., Tabellenband und Atlas.
- ESRI (2020): ESRI official website, accesses 01 August 2021, <https://www.esri.com/en-us/home>.
- Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I., & G. Tutz (2016): Statistik: Der Weg zur Datenanalyse (8th ed.). Wiesbaden, Germany: Springer, p. 581.
- Faulhaber, P. (1996): Flußbauliche Analyse und Bewertung der Erosionsstrecke der Elbe. In: Mitteilungsblatt der Bundesanstalt für Wasserbau, 74:33–49.
- FFG Elbe [Flussgebietsgemeinschaft Elbe] (2020): Entwurf der zweiten Aktualisierung des Bewirtschaftungsplans nach § 83 WHG bzw. Artikel 13 der Richtlinie 2000/60/EG für den deutschen Teil der Flussgebietseinheit Elbe für den Zeitraum von 2022 bis 2027., accessed 01 August 2021, https://beteiligung.fgg-elbe.de/bp/PDF-Anlagen/Entwurf_BP_Stand_12-2020.pdf
- FFG Elbe [Flussgebietsgemeinschaft Elbe] (2021): The Elbe River Basin, accessed 01 August 2021, <https://www.fgg-elbe.de/index.php/einzugsgebiet-en.html>
- Gabriel, T., Kühne, E., Kloth, S., Anlauf, A., Götz, E., Faulhaber, P., Puhlmann, G. & T. Gröger. (2009): Sohlstabilisierungskonzept für die Elbe – von Mühlberg bis zur Saalemündung. Wasser- und Schifffahrtsverwaltung (Eds.), Magdeburg, Dresden, Koblenz, Karlsruhe, 2009, p. 102.
- Gilbert, G. W. (1914): The Transportation of Debris by Running Water, US Geological Survey Professional Paper 86, Washington, DC: US Government Printing Office.
- GeoBasis-DE/BKG (2021): Digitales Basis-Landschaftsmodell (Ebenen) (Basis-DLM), Bundesamt für Kartographie und Geodäsie (Eds.), accessed 01 August 2021, <https://gdz.bkg.bund.de/index.php/default/digitales-basis-landschaftsmodell-ebenen-basis-dlm-ebenen.html>.
- Gómez-Rubio, V. (2020): Bayesian Inference with INLA. Chapman & Hall/CRC Press. Boca Raton, FL.
- GRASS Development Team (2021): GRASS GIS 7.8.6dev Reference Manual, accessed 01 August 2021, <https://grass.osgeo.org/grass78/manuals/helptext.html>
- Haunschild, A. (1996): Zur Sohlstruktur der Elbe. In: Prange, A. (Eds.) 1996: Ökosystem Elbe – Zustand, Entwicklung und Nutzung: 7. Magdeburger Gewässerschutzseminar; Internationale Fachtagung in Budweis, Tschech. Republik, vom 22. - 25. Okt. 1996, Budweis.
- Hijmans, R. (2021): Spatial autocorrelation. In: Spatial Data Science with R, accessed 01 August 2021, <https://rspatial.org/raster/index.html>.

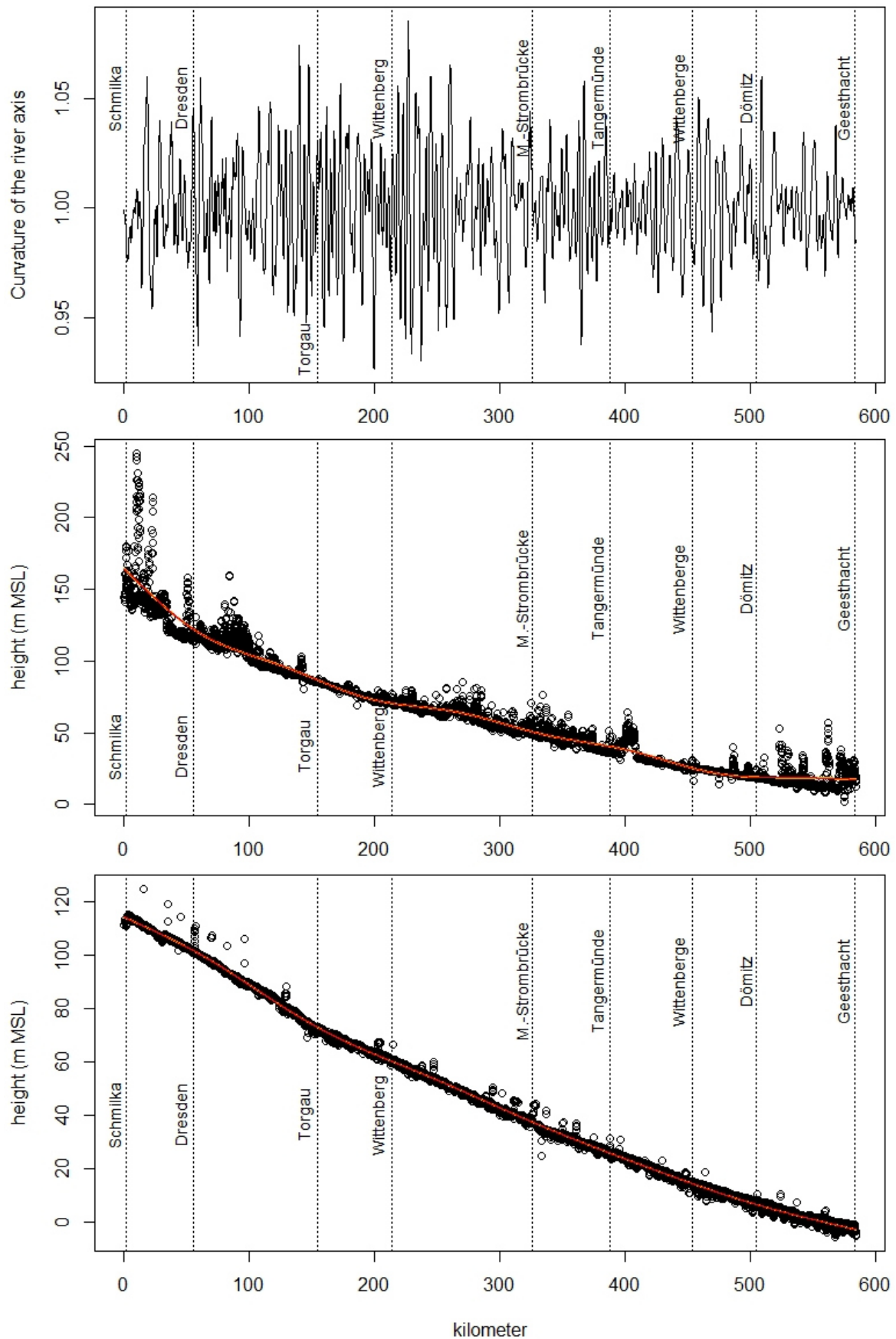
- Mannsfeld, K. & A. Bernhardt (2008): Sächsische Schweiz. In: *Naturräume in Sachsen. Forschungen zur deutschen Landeskunde*, 257. Mannsfeld, K. & Syrbe, R.-U. (Eds.), Deutsche Akademie für Landeskunde, Selbstverlag, Leipzig, p. 288.
- March, D., Alós, J., Cabanella-Reboredo, M., Infantes, E., Jordi, A. & M. Palmer (2013): A Bayesian spatial approach for predicting seagrass occurrence. In: *Estuarine, Coastal and Shelf Science* 2013, 131:206–212.
- Martins, T. G., Simpson, D., Lindgren, F., & H. Rue (2013): Bayesian computing with INLA: New features. In: *Computational Statistics & Data Analysis*, 67:68–83.
- Meyer, T. (2017): *Ökologie mitteleuropäischer Flussauen*. Springer Verlag, Berlin, Germany, p. 163.
- Mitra, S. (2009): Spatial Autocorrelation and Bayesian Spatial Statistical Method for Analyzing Intersections Prone to Injury Crashes. In: *Transportation Research Record: Journal of the Transportation Research Board*, Transportation Research Board of the National Academies, Washington, D.C., 2009, 2136:92–100.
- Mizumoto, A. & L. Plonsky (2016): R as a lingua Franca: Advantages of using R for quantitative research in applied linguistics. *Applied Linguistics*, 37,2:284–291.
- Moraga, Paula. (2019): *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. Chapman & Hall/CRC Biostatistics Series, p. 274.
- Nestmann, F & Büchele, B. (2002): Projektüberblick. In: *Morphodynamik der Elbe. Schlussbericht des BMBF-Verbundprojektes mit Einzelbeiträgen der Partner und Anlagen-CD*. Nestmann, F & Büchele, B. (Eds.), Institut für Wasserwirtschaft und Kulturtechnik Karlsruhe (TH) 2002, Karlsruhe.
- Neteler, M. (2001): Kurzbeschreibung von GRASS GIS, accessed 01 August 2021, https://grass.osgeo.org/gdp/tutorial/grass_shortintro_de.html.
- Redding, D. W., Lucas, T. D. C., Blackburn, T. M. & K. E. Jones (2017): Evaluating Bayesian spatial methods for modelling species distributions with clumped and restricted occurrence data, *PloS one*, 12(11).
- R-INLA (2021): R-INLA Project, accessed 01 August 2021, <https://www.r-inla.org/home>.
- Rommel, J. (2000): Karten zur Laufentwicklung der Elbe. Mittlere Elbe um 1730 –1780 – 1830 – 1890. Studie im Auftrag der Bundesanstalt für Gewässerkunde, Koblenz: 1-61, accessed 01 August 2021, <http://elise.bafg.de/>.
- RStudio Team (2021): RStudio. RStudio, PBC accessed 01 August 2021, <https://www.rstudio.com/products/rstudio/>.
- Rue, H., Martino, S. & N. Chopin (2009): Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. In: *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 71,2:319–392.
- Saucke, U. & Brauns, J. (2002): Stromtalgeschichte und Flussgeologie im deutschen Elbegebiet. In: *Morphodynamik der Elbe. Schlussbericht des BMBF-Verbundprojektes mit Einzelbeiträgen der Partner und Anlagen-CD*. Nestmann, F & Büchele, B. (Eds.), Institut für Wasserwirtschaft und Kulturtechnik Karlsruhe (TH) 2002, Karlsruhe.
- Saucke, U., J. Rommel, and J. Brauns. (1999): Die Geologie Der Elbe. In: *Morphodynamik der Elbe. Schlussbericht des BMBF-Verbundprojektes mit Einzelbeiträgen der Partner und Anlagen-CD*. Nestmann, F & Büchele, B. (Eds.), Institut für Wasserwirtschaft und Kulturtechnik Karlsruhe (TH) 2002, Karlsruhe.
- Schenk, O. (2021): PARDISO, PARDISO 7.2 Solver Project (December 2020), Institue of Computing, USI Lugano, accessed 01 August 2021, <https://pardiso-project.org/>.
- Schindler, S., O'Neill, F. H., Biró, M., Damm, C., Gasso, V., Kanka, R., van der Sluis, T., Krug, A., Lauwaars, S. G., Sebesvari, Z., Pusch, M., Baranovsky, B., Ehlert, T., Neukirchen, B., Martin, J. R., Euller, K., Mauerhofer, V. & T.Wrbka (2016): Multifunctional floodplain management and

- biodiversity effects: a knowledge synthesis for six European countries. In: *Biodiversity and conservation*, 25,7:1349–1382.
- Schmidt, A. & A. Boyd (2020): *R Software Handbook. Evaluation, Statistics, and Methodology* – University of Tennessee, Knoxville.
- Simon, M. (2018): *Die Elbe im Raum Magdeburg. Darstellung der wasserwirtschaftlichen Verhältnisse*. Landesbetrieb für Hochwasserschutz und Wasserwirtschaft Sachsen-Anhalt (Eds.), accessed 01 August 2021, https://lhw.sachsen-anhalt.de/fileadmin/Bibliothek/Politik_und_Verwaltung/Landesbetriebe/LHW/neu_PDF/1.0/Z_Fraebel_1-LHW_Oeffentlichkeitsarbeit_Broschuere_Herr_Simon_Si_Die_Elbe_im_Raum_Magdeburg_komprimiert.pdf.
- Taylor, B. M. & P. J. Diggle (2012): INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. In: *Journal of statistical computation and simulation*, 84,10:2266–2284.
- Tockner, K., & J. Stanford (2002): Riverine flood plains: Present state and future trends. In: *Environmental Conservation*, 29,3:308–330.
- Tschirk, W. (2014): *Statistik: Klassisch oder Bayes: Zwei Wege im Vergleich* (2014th ed.). Wiesbaden, Germany: Springer, p. 166.
- Turkman, A. M. A., Paulino, C. D., & P. Müller (2019): *Computational Bayesian Statistics*. Cambridge University Press.
- Umweltamt Landeshauptstadt Dresden (2010): *Gewässersteckbrief Elbe*. Umweltamt Landeshauptstadt Dresden, 2010, accessed 01 August 2021, <http://www.dresden.de/media/pdf/umwelt/umweltausstellung/Elbe.pdf>.
- van de Schoot, R., Depaoli, S., King, R., Kramer, B., Märtens, K., Tadesse, M. G. & C. Yau (2021): Bayesian statistics and modelling. In: *Nature Reviews Methods Primers*, 1(1).
- van Niekerk, J., Bakka, H., Rue, H. & O. Schenk (2019): New frontiers in Bayesian modelling using the INLA package in R. In: *Journal of Statistical Software*, accepted, in press.
- Ward, J. V. (1998): Riverine landscapes: Biodiversity patterns, disturbance regimes, and aquatic conservation. In: *Biological Conservation*, 83,3:269–278.
- Ward, J. V., Tockner, K., Arscott, D. B., & C. Claret. (2002): Riverine Landscape Diversity. In: *Freshwater Biology* 47:517–539.
- Weber, A. (2020): Digital elevation model (DEM 1) of the River Elbe floodplain between Schmilka and Geesthacht, Germany [Data set]. PANGAEA – Data Publisher for Earth & Environmental Science.
- Weber, A., & M. Hatz (2020): Hyd1d – Algorithms to Compute 1D Water Levels Along German Federal Waterways Elbe and Rhine. In: *Environmental Modelling & Software* 1,1: 1–1.
- Weber, A., & S. Rosenzweig (2020): hyd1d - Algorithms to compute flood extent and duration along German federal waterways Elbe and Rhine. In: *Journal of Ecohydraulics*, 1,1:1–1.
- Witt, J. L. (1995): Introduction. In: *Sharing the challenge: Floodplain management into the 21st century: Interagency floodplain management review committee of the floodplain management task force*. *Environment*, 37,1:25–28.
- WSV [Wasserstraßen- und Schifffahrtsverwaltung des Bundes] (2020): *Verkehrsnetz Bundeswasserstraßen (VerkNet-BWaStr)*, accessed 01 August 2021, https://www.gdws.wsv.bund.de/DE/service/karten/03_VerkNet-BWaStr/VerkNet-BWaStr_node.html.
- WWF Deutschland (2021): *Mittlere Elbe – Mehr Raum für Fluss und Aue*, accessed 01 August 2021, <https://www.wwf.de/themen-projekte/projektregionen/elbe>.

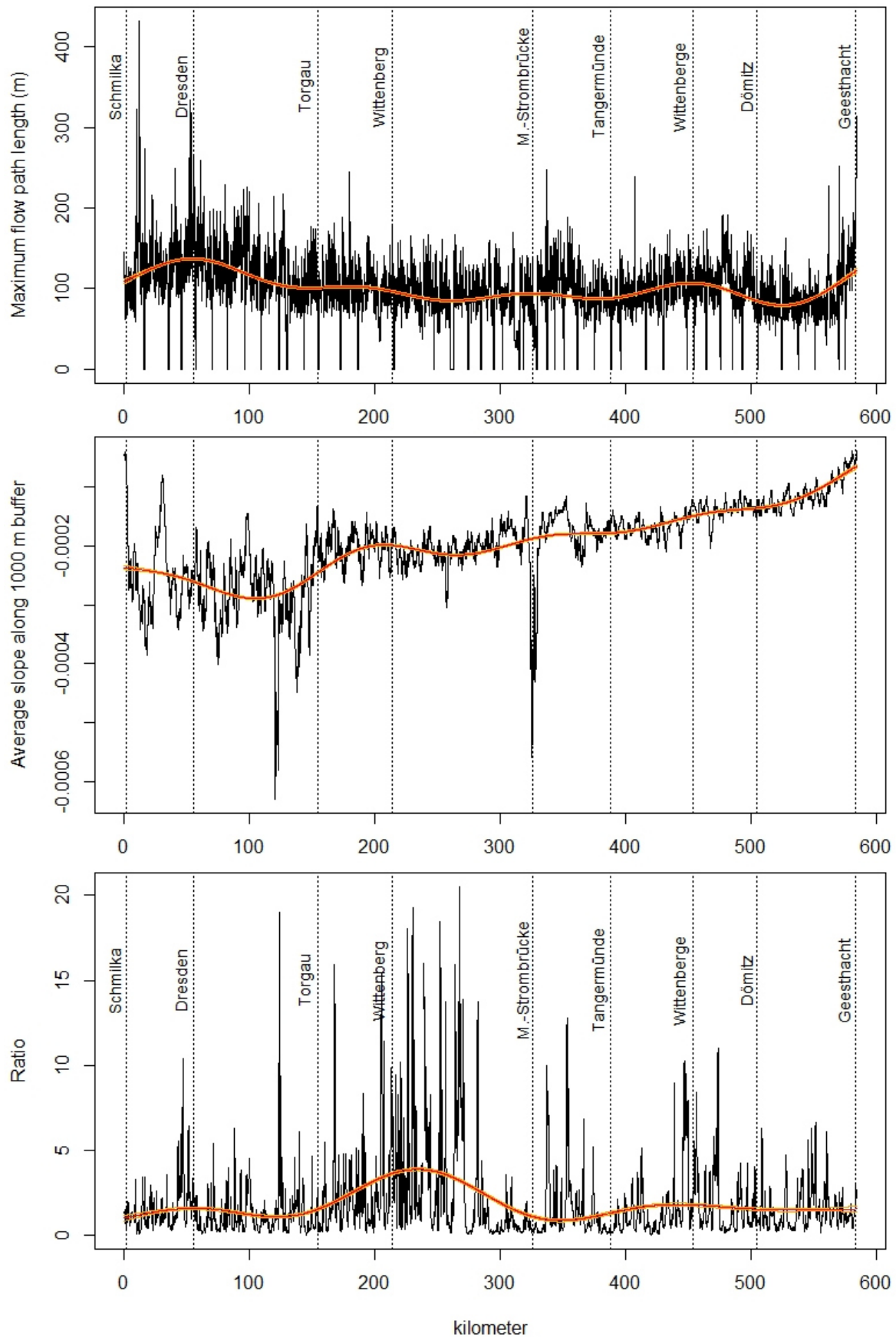
Appendix



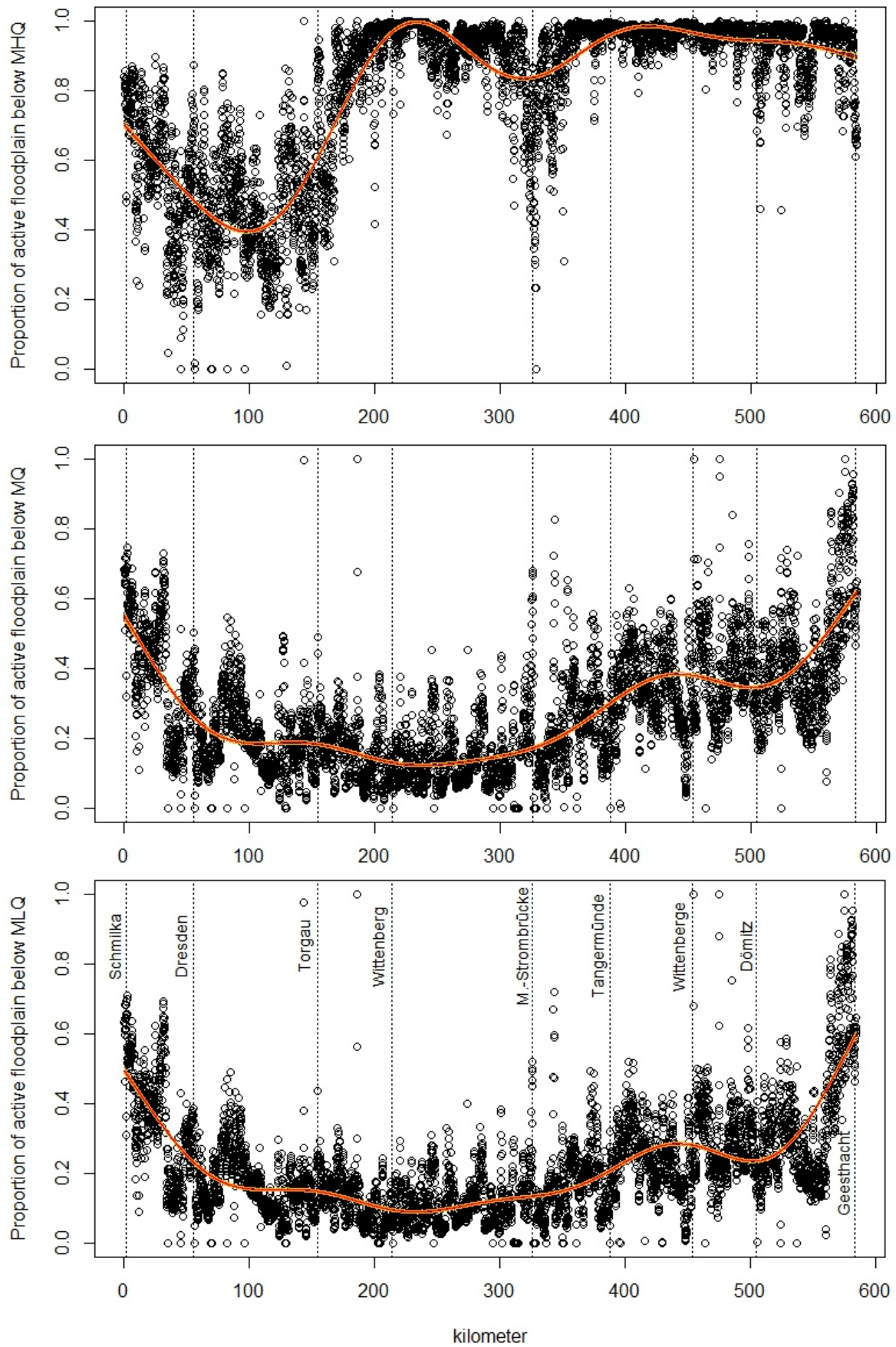
Appendix 1 Top: The returns of the function `getAngle()`, middle: the returns of the function `getAxisADiff()`, bottom: the returns of the function `getAxisPDiff()`. All returns are plotted against the Elbe kilometrage. Additionally, a gam regression curve with upper and lower standard deviation is depicted in the bottom two plots.



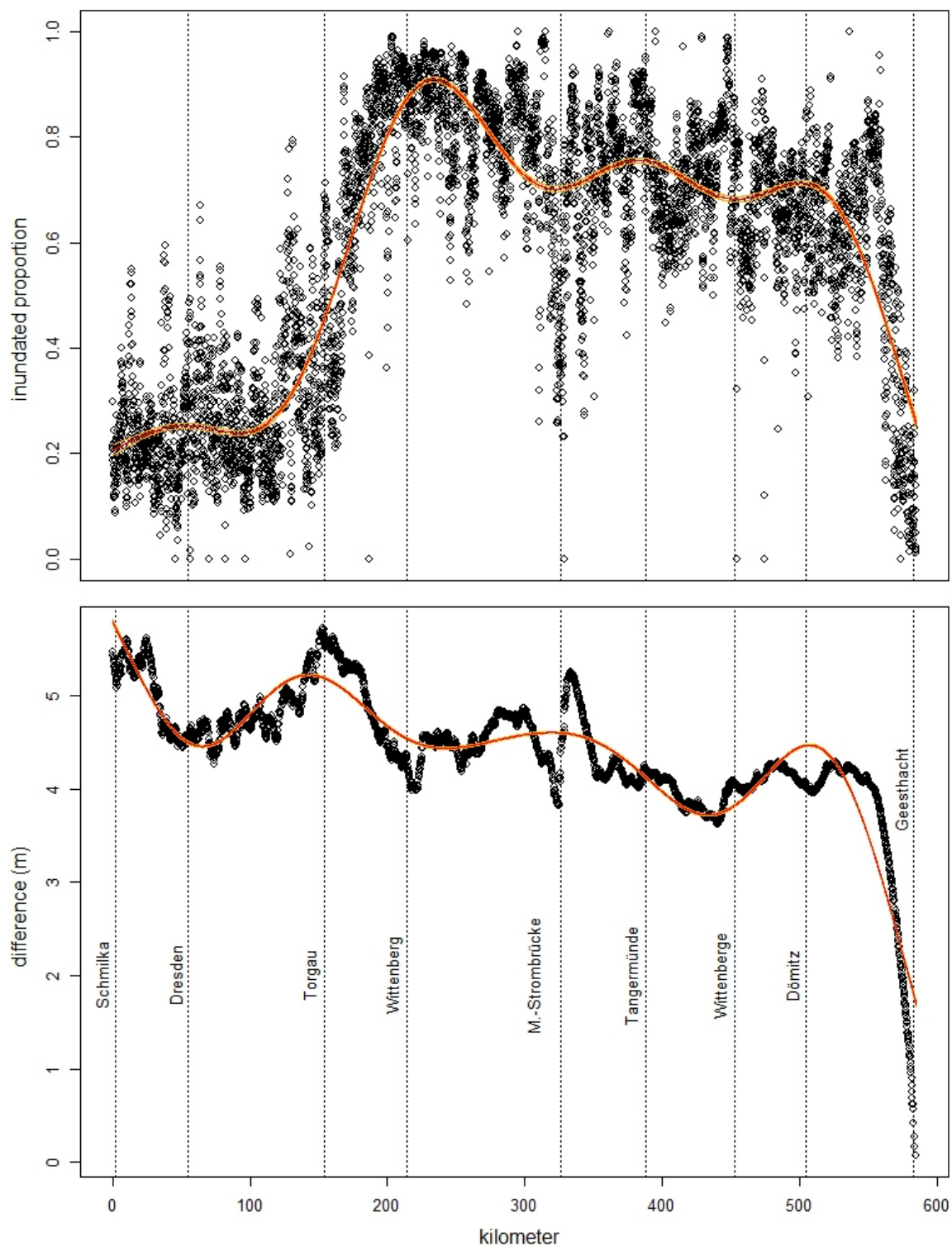
Appendix 2 Top: The returns of the function *getCurve()*, middle: the returns of the function *getDEMmax()*, bottom: the returns of the function *getDEMmin()*. All returns are plotted against the Elbe kilometrage. Additionally, a gam regression curve with upper and lower standard deviation is depicted in the bottom two plots.



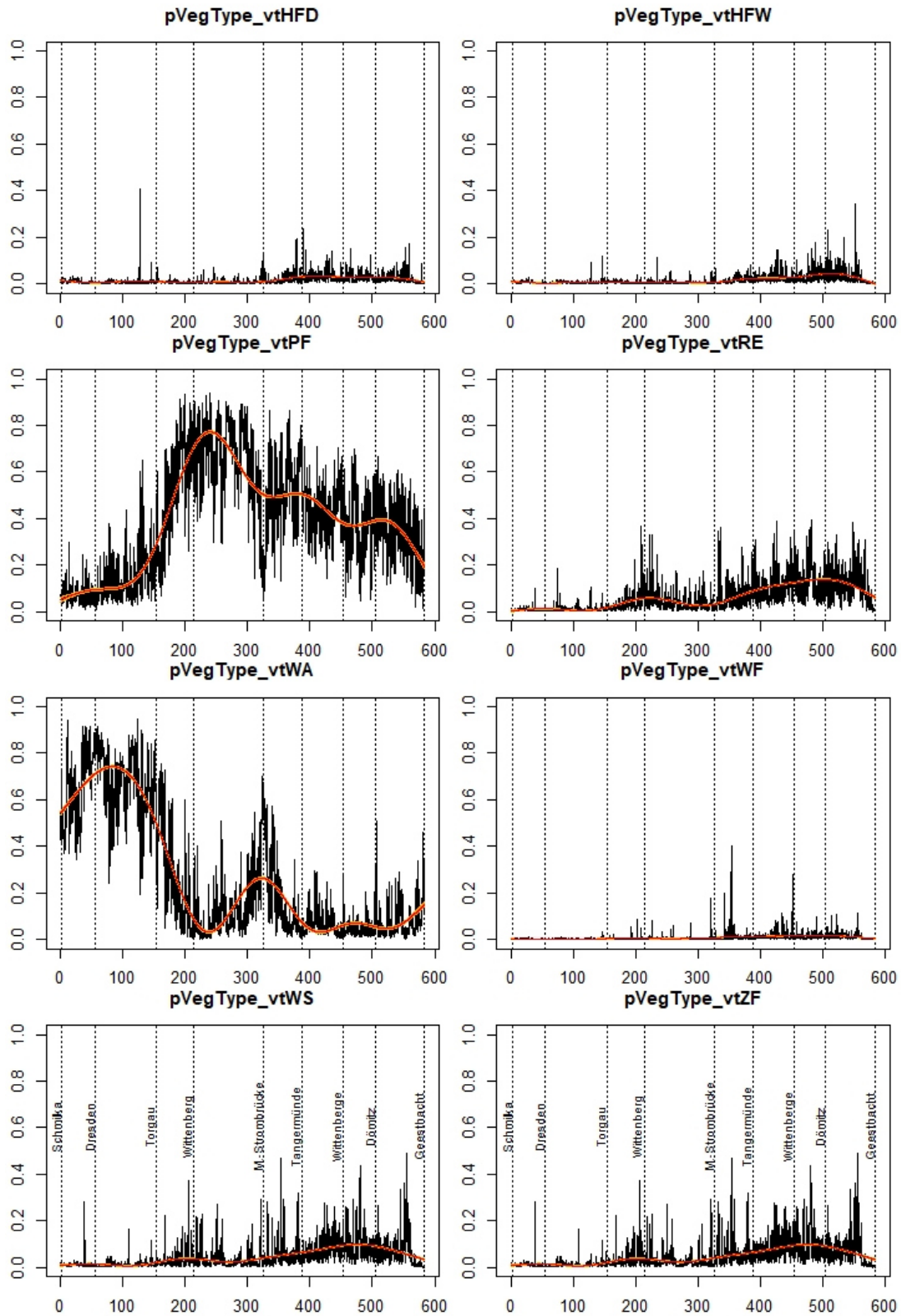
Appendix 3 Top: The returns of the function `getMaxFlowLength()`, middle: the returns of the function `getSlope()` with 1000 m buffer, bottom: the returns of the function `getRatio()`. All returns are plotted against the Elbe kilometrage. Additionally, a gam regression curve with upper and lower standard deviation is depicted.



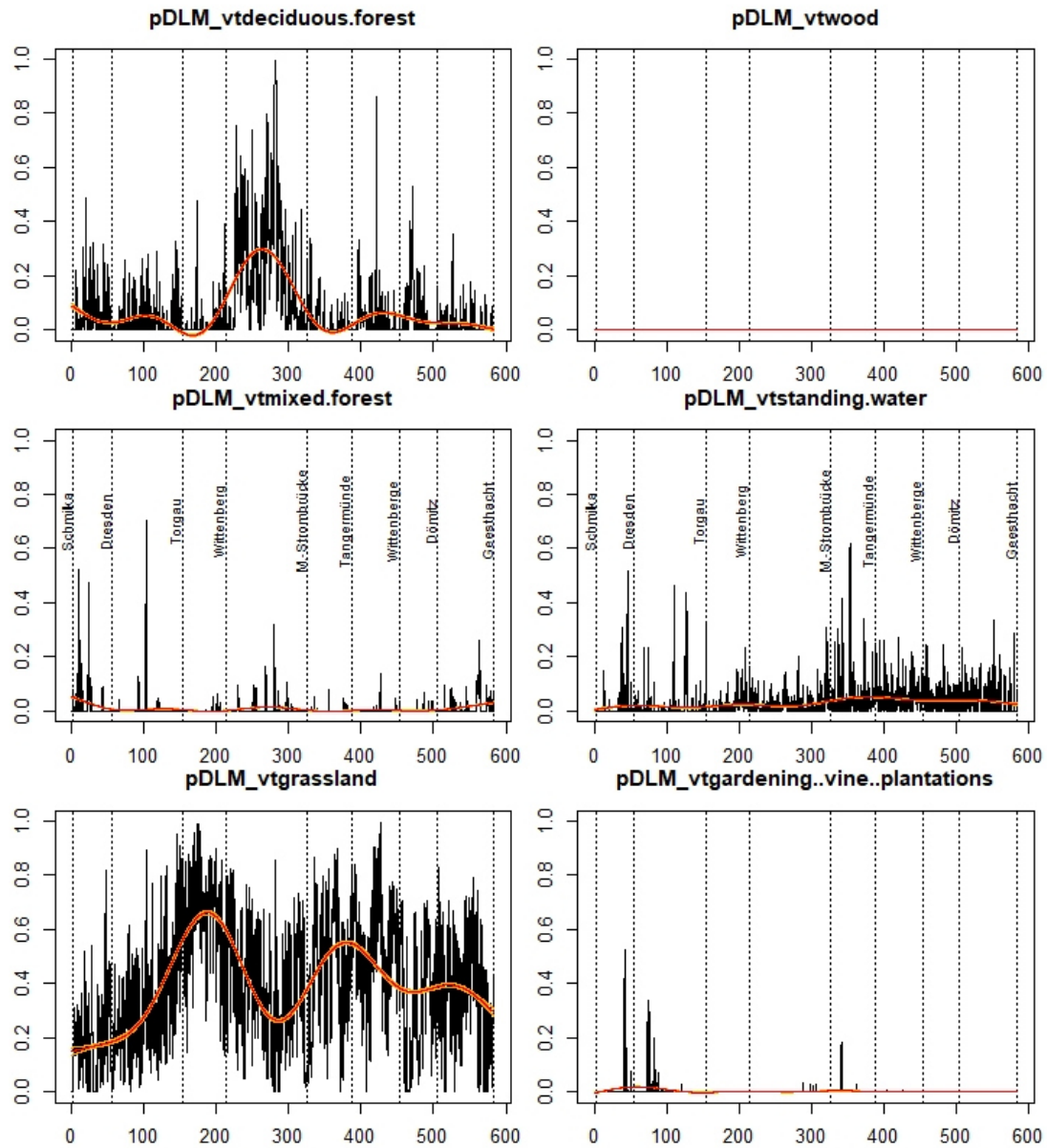
Appendix 4 Top: The returns of the function *getPbelowMHQ()*, middle: the returns of the function *getPbelowMQ()*, bottom: the returns of the function *getPbelowMLQ()*. All returns are plotted against the Elbe kilometerage. Additionally, a gam regression curve with upper and lower standard deviation is depicted.



Appendix 5 Top: The returns of the function `getPbetweenMLQMHQ`, bottom: The returns of the function `getDiffbetweenMNQMHQ`. All returns are plotted against the Elbe kilometerage. Additionally, a gam regression curve with upper and lower standard deviation is depicted.



Appendix 6 The returns of the function `getVegtype()`. The different queried PNV types are depicted as their proportion of the active floodplain. All returns are plotted against the Elbe kilometrage. Additionally, a gam regression curve with upper and lower standard deviation is depicted.



Appendix 7 The returns of the function `getDLM()`. The different queried DLM25 types are depicted as their proportion of the active floodplain. All returns are plotted against the Elbe kilometrage. Additionally, a gam regression curve with upper and lower standard deviation is depicted.