UNIVERSITY of GLASGOW

P-spline Additive Modeling and Partial Derivative Estimation for Environmental Data

by

George Vazanellis

A thesis submitted in partial fulfillment for the
degree of Doctor of Philosophy

in the

School of Mathemetics and Statistics
College of Science and Engineering

December 2019

Declaration of Authorship

I, George Vazanellis, declare that this thesis titled, 'P-spline Additive Modelling and Partial Derivative Estimation for Environmental Data' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

"You're worried about loosing your hair? Just make sure you don't loose your teeth."

Nick Vazanellis

UNIVERSITY of GLASGOW

Abstract

School of Mathemetics and Statistics
College of Science and Engineering

Doctor of Philosophy

by George Vazanellis

This thesis addresses the construction of complex additive mixed models for environmental data and the use of those models to estimate partial derivatives for the purpose of detecting impacts of known events.

The methods developed are applied to a data set collected by the Scottish Environment Protection Agency in an effort to monitor the dissolved oxygen of the River Clyde. There are many metrics recorded along the River. Exploratory analysis is carried out to pinpoint some possible drivers of the dissolved oxygen.

The River Clyde contains processes which are difficult to represent by conventional parametric models. P-splines offer a means of fitting a flexible model to this data set. There is also the possibility of the presence of interactions between some explanatory covariates. Because of the sampling regime, a random effects component is appropriate. An additive mixed model with interactions allows for all the above-mentioned components to be included in a representative model for the River Run data. The methodology for fitting such a model, along with descriptions of four information criteria which are intended to aid in smoothing parameter selection, are explained in this thesis. Two options for performing analysis of variance for additive models with interactions are considered: A simple F-test and a quadratic approach. The performance and computational expense of each is compared to a parametric bootstrap and to various other standard tests.

A simple additive model with no interactions is initially fitted with varying degrees of freedom for each main effect. The four information criteria scores are calculated for every main effect across all degrees of freedom. The information criterion which performs best is then used to select the optimal smoothing parameter for every main effect in an additive model and an additive mixed model, both with no interactions. Before an additive mixed model with interactions is fitted, a simulation study is conducted to see if the order of optimization of the main effect degrees of freedom is of any importance. An additive mixed model with interactions is subsequently fitted and interpreted.

One aim of this thesis is to determine if upgrades to two wastewater treatment facilities have had positive impacts to the levels of dissolve oxygen in the river. Partial derivatives with respect to time are discussed as a means of detecting subtle changes in a system which has shown gradual increases in dissolved oxygen over the past four decades. An argument is made for the use of P-splines with penalty orders other than 2 if the main goal is derivative estimation. A simulation study is conducted and the optimal penalty order is then used to construct a derivative additive mixed model with interactions for the River Run data. This model is used to see if there is evidence the wastewater facility upgrades had a positive impact. One positive result of this research is that the quadratic forms method of analysis of variance for additive models with interactions

was found to out-perform the simple F-test and was less computationally expensive than the parametric bootstrap. A second positive result was finding a preferred information criterion for smoothing parameter selection and using the optimal degrees of freedom to subsequently fit such a complex additive mixed model with interactions. A third positive result was finding that penalty order three outperformed penalty order two in estimating partial derivatives. Finally, the fourth positive result was constructing a derivative model and subsequently using it to provide evidence the wastewater treatment facility upgrades had a positive impact on the dissolved oxygen.

Keywords: P-splines, additive mixed model, derivative estimation, penalty order, environmental data

Acknowledgements

I want to thank my advisor, Professor Adrian Bowman, for his unbounded support, patience, and kindness. I also want to thank his wonderful family for opening their house and showing me so much hospitality. Adrian has treated me like a friend and a brother when I needed it most. God bless that man. I also want to thank all the faculty and staff of the Department of Mathematics and Statistics of the University of Glasgow for their support and compassion.

Sometimes students neglect themselves while pursuing a PhD. Ruth and Shazia made sure to remind me to eat properly, get enough sleep, and to keep on task. For this I cant thank them enough. I could not ask for better colleagues or friends.

I will never forget all the wonderful officemates I was fortunate enough to have over the years. Thank you Cunyi, Marnie, Irena, Craig, Umberto, Diana, Amira, Menyi, Qing, Anna, Alan, Fluke, Miriam, Ivona, Michael, and Ben for your support and friendship. A special thanks goes to Craig Wilke, Benn McDonald, and Caroline Haig for answering any questions I would ask and for being so nice about it.

I want to thank the SECURE Network for granting me a feasibility project grant allowing me to work with the Scottish Environment Protection Agency over a six month period. Dr Alan Hills, Dr Ted Schlicke, and the whole OceanMod team offered their invaluable expertise and support. I appreciate all their efforts.

Michael, Louie, and Komis are my dearest friends. They helped me get through this journey. These were the people I would call when I was doubting my ability to achieve this accomplishment. They would always prop me up and convince me I could do it. I love you guys.

The biggest debt of gratitude goes to my mother Vera, my father Nick, my nephew Victor, my sister Sly, and my niece Veroula. Thank you for all your love and support. I love you all very much. Veroula, I want to thank you on behalf of the rest of our family for taking care of all of us. You are a star!

Finally, I want to thank Mina for inspiring me to pursue a lifelong dream and providing the motivation to see it through. Thank you for taking care of our beautiful son. I love you guys.

Contents

List of Figures

List of Tables

Dedicated to Nikolaki, Armando, and to the memory of my dear friend Michael.

Chapter 1

Introduction

The dissolved oxygen in a body of water is essential to sustain most forms of life contained in it. A substantial drop in the concentration of dissolved oxygen can have devastating results such as large amounts of fish being killed. Knowing the conditions which may cause these drops can allow those concerned to take specific measures to minimize the damage. The River Clyde is the body of water considered in this thesis.

The River Clyde has been and continues to be monitored by the Scottish Environment Protection Agency (SEPA). The result of this monitoring is the River Run data set spanning over forty years. A statistical model generated from the data can provide valuable insight on the drivers of the dissolved oxygen. One application of a representative model can be its use as a forecasting tool. There is great benefit in knowing not only the individual detrimental conditions, but also what combination of conditions pose a higher threat. A second application can be the use of the model's capacity to capture the river's history. The subtle effects of past events can be identified and can aid in future decision making and allow for possible intervention averting ecological disasters.

The River Clyde represents a complex system and it would be naive to expect a simple statistical model would truly be representative. Physical and chemical processes both have some influence on the amount of dissolved oxygen in any body of water. Metrics of these reproducible phenomena are ideal candidates to be represented by main effects of a model. Furthermore it may be necessary to allow some of these fixed effects to have an increased degree of flexibility by introducing smoothed terms. One reason may be biological processes in place may not allow certain physical and/or chemical effects to behave in a simple or conventional manner.

P-splines are a relatively new approach to fitting additive models and the River Run data is an ideal data set for these methods because of the several periodic explanatory

covariates, such as the tides and day of year. Also, the year and station effects on disolved oxygen are not expected to behave in a simple parametric manner. Additive models are widely used to describe systems with complex relationships between the covariates and the response. These will be used in the thesis, using the popular P-spline formulation.

Considering not all influences on the dissolved oxygen are represented in the data, it stands to reason conditions in place on different sampling dates may account for a substantial proportion of the variability of the dissolved oxygen. Inclusion of random effects terms will allow for the explanation of any increased variability. Therefore, an additive mixed model will allow for any needed flexibility, allow for the inclusion of multiple covariate interaction terms, and account for any random effects.

Detecting the impact of specific events in the environment is a very important activity. The Clyde data is an important example. The dissolved oxygen levels of the River Clyde have generally been increasing during the past four decades. There is however interest in determining whether abrupt yet subtle increases in the dissolved oxygen can be attributed to known events such as wastewater facility upgrades. The presence of a step increase may be difficult to detect by simply looking at the time main effect of an additive model.

Many methods for detecting change assume that the system is in some kind of steady-state from which deviations can be identified. However, it is very common to be in a situation where the process under study is dynamic and changing all the time. In particular, the Clyde data show continual increase in water quality. This makes the detection of additional positive impact of event more challenging. Estimation of partial derivatives over time offers a means of focusing on periods of particularly rapid change. The aim is to deliver an analysis of the Clyde data which delivers real insight into the effects of sewage treatment upgrades. This will be discussed in the context of P-spline additive models and methods of fitting these models, including a suitable choice of penalty, will be investigated.

Graphical tools will be developed to facilitate the visualization of the behavior of complex multivariate interaction terms. Plots depicting the behavior of the model main effect and interaction terms will be used to allow one to get a sense of the type of influence on the dissolved oxygen attributed to each fixed effect and combinations of fixed effects. These tools will also help in the detection by means of partial derivatives.

This thesis is comprised of 8 chapters. Chapter 1 presents the motivation of this thesis; there is value in fitting an additive model with interactions and random effects to describe the dissolved oxygen of a system as complex as the River Clyde. Furthermore, the use

of these models as tools for partial derivative estimation to aid in the detection of effects brought on by known effects is proposed. The graphical representations of these partial derivative estimates will play a large role in the detection of these known events. Furthermore, investigation of the effect of various penalty orders of terms in the models which will be used as derivative estimators is proposed.

Chapter 2 describes the River Run data. Detailed tables will list all the variables in the data set along with any transformations producing other appropriate variables. Boxplots and bar-charts depicting the distribution of the variables will be presented with a summary of any issues in the data collection. Also, scatterplots of dissolved oxygen against all potential covariates are presented to explore any underlying relationships of the response variable with the explanatory variables. Subsequently, explanatory variables plotted against each other to depict any relationships between the covariates are presented and discussed. Time series plots of the dissolved oxygen with appropriate marginal smooths superimposed for each station are also presented and discussed.

Chapter 3 describes the methodology to be used for constructing an additive mixed model. A detailed summary of B-splines is presented. This summary begins with the definition of a single B-spline of degree 0, with this definition used to construct a B-spline of order q by iteration. Subsequently, the construction of a B-spline basis is outlined. These concepts are extended from B-spline functions of a single variable to functions of two and three variables. What follows is the description and implementation of a penalty term to create a P-spline model for the univariate, bivariate, and trivariate cases. The construction of additive mixed models containing univariate, bivariate, and trivariate terms requires identifiablility constraints. An algorithm for the implementation of a random effect is also required. These issues are discussed in detail. Chapter 3 closes with the details of various smoothing parameter selection criteria.

Chapter 4 addresses analysis of variance of additive model terms as a means of additive model selection, through a simulation study. The chapter begins by outlining the type of additive model which is to be analyzed. Three methods of computing the required distribution of the F-statistic (simple F test, quadratic forms, and parametric bootstrap) are explained. Subsequently, the design of the simulation study is described. Bivariate and trivariate functions, each containing a bivariate interaction term, generate outputs with random noise added. Initially, the effect of the interaction term is nullified. The effect of the interaction term is then incrementally increased to see how the three methods perform in detecting the presence of the effect. This is done by plotting the proportion of P-values under 0.05 against the effect size.

Chapter 5 details the construction of an additive mixed model with interactions for the River Run data. Initially a simple additive model, one which is comprised of only univariate main effects, is fitted, with a default of 6 degrees of freedom for each main effect. The same model is then fitted with a range of degrees of freedom and the smoothing parameter selection criteria discussed in Chapter 3 are compared. The selected criterion is then used to determine appropriate degrees of freedom for the main effects of a simple additive model and a simple additive mixed model. A simulation study is conducted to see if the degrees of freedom optimization order is of importance when fitting a bivariate additive model with interactions. A complex additive mixed model, with main effects, bivariate interaction terms, trivariate interaction terms, and random effects, is fitted with the optimal smoothing parameter selection criteria used to determine the degrees of freedom for each term. Several permutations of the order of degrees of freedom optimization are tried and the results compared. Plots of the main effects and interaction terms are presented and interpreted for all models fitted throughout the chapter. The final additive mixed model with interactions is analyzed and discussed

Chapter 6 addresses derivative and partial derivative estimation via P-spline additive models. The chapter begins with a literature review of general derivative estimation and delves deeper into the literature specifically discussing derivative estimation using P-splines. Although Eilers and Marx (1996) claimed there was nothing special about the second derivative penalty, and thus any penalty order could be used, justification for penalty order 3 and 2/3 is given for derivative and partial derivative estimation. Chapter 6 goes on to describe the design of a simulation study to assess the optimal penalty order for derivative and partial derivative estimation. Various univariate and bivariate functions with their derivatives and partial derivatives respectively are graphically depicted. The simulation study is then carried out with varying penalty orders and box-plots of the root mean squared error distribution are presented for each function across all penalty orders. Bar-charts of the overall performance of the penalty orders are then presented.

Chapter 7 proceeds to fit a derivative additive mixed model with interactions for the River Run data and used this to detect the impact of upgrades of two wastewater treatment facilities on the dissolved oxygen levels. The derivative model is fitted in the manner outlined in Chapter 5, with the optimal penalty criterion found in Chapter 6. Plots of the main effects and interactions are presented with a special focus on the trivariate interactions in close proximity to the upgraded treatment facilities.

Chapter 8 summarizes the findings of this research, highlights its novelty, and suggest future work to extend the research.

Chapter 2

Clyde Data

2.1 A Brief Description and History of the River Clyde

The Clyde estuary receives organic waste from the large conurbation of the Greater Glasgow region. Degradation of this organic matter deposited in the sediments and suspended in overlying water meant that the inner estuary was essentially devoid of dissolved oxygen in the 1970s, although there were slightly higher concentrations in the outer estuary. Dissolved oxygen concentrations increased as discharges of organic waste declined as a result of improvements to effluent treatment. This increase in dissolved oxygen resulted in the reappearance of salmon in 1983 following an absence of over a century. Salmon have reappeared each year since, along with increasing numbers of other fish species.

Further reductions in discharges of organic wastes may not result in an immediate increase in dissolved oxygen as organic matter trapped in the sediments continues to remove oxygen from the overlying water. Land reclamation and dredging have made the inner Clyde narrower and deeper than its natural state. These changes have reduced the strength of the tidal currents causing less mixing between freshwater and the incoming seawater, resulting in layering of the water column. This layering inhibits the transfer of oxygen from the surface allowing deeper water to become depleted in oxygen. A computer program which simulates the intricate flows of the Clyde estuary is being used to predict the most cost effective option for increasing dissolved oxygen in the Clyde. These options include:

- Further reducing inputs of organic waste.
- Changing the operation of the tidal weir to increase mixing between fresh and salt water.

- Injecting pure oxygen into the estuary at critical times (e.g. during extended dry, hot spells) and locations.
- Changing the morphology of the estuary to improve flushing and reduce the residence time of organic waste

This background information is taken from Baxter et al. (2011) and Kielmas (2019) and can be found using the link

http://marine.gov.scot/datafiles/misc/MarineAtlas-Complete.pdf

2.2 Describing the SEPA Data Set

As *P*-spline additive modeling and partial derivative estimation for environmental data is the focus of my PhD research, I am fortunate to have access to a rich data set compiled by the Scottish Environment Protection Agency (SEPA). Furthermore, my time spent at SEPA offices during a six month secondment provided essential insight into the nature of the Clyde Estuary and all the known drivers of the oxygen levels. This was made possible through one of the SECURE Network (Statistics of Environmental Change, Resources, and Ecosystems) Feasibility Projects. The data set involved in this research is the River Run data.

2.2.1 Collecting the River Run Data

SEPA has been collecting data from various points in the Clyde estuary for the purpose of monitoring the state of river's health. Samples of river water are collected by SEPA personnel by boat and hence the resulting data set has come to be referred to as the "River Run" data. The data used in this research includes observations collected during 669 sample surveys from January 1970 to December 2015. SEPA continues to conduct River Run sampling to this day and all analysis presented within this research can be applied to any augmented River Run data.

There are fifteen distinct sampling locations of the River Run as depicted in Figure 2.1. Easting and northing variables contained in the River Run data are used to pinpoint these sampling locations. The majority of the data were collected from the thirteen sampling points shown in red. These locations are separated by approximately 2 mile intervals from Broomielaw in Glasgow going west to Kempock Point in Gourock. The blue points represent the sampling locations less frequented than the sampling locations represented by the red points. The multiple sampling locations allow one to include

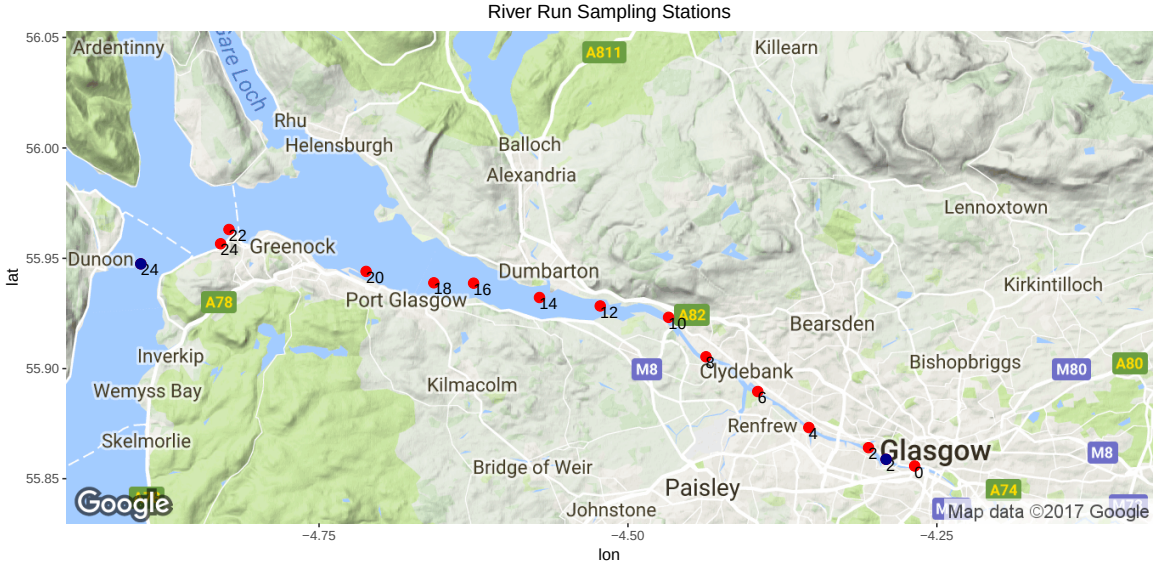


FIGURE 2.1: The above map shows the locations of the sampling stations with the distance from weir values in miles for the Clyde river run data. The red dots represent the sites sampled most frequently and the blue dots represent the less frequent sampling sites. The blue dots were given distance from weir values of 2 and 24 miles.

a spatial component to a statistical model(McMullan et al. (2003)). Mile 24 in blue represents the sampling location described in the data as "Firth of Clyde at CMT 7, NW of Cloch Point" which was only sampled in 1999. Mile 24 in red represents the sampling location described in the data as "Clyde Estuary at Kempock Point, Gourock" which was sampled on all other occasions of data collection. Mile 2 in red represents the sampling location described in the data as "Clyde Estuary @ Kelvin Confluence" which was collected throughout the river run. Mile 2 in blue represents the sampling location described in the data as "CLYDE ESTUARY AT BELL'S BRIDGE, Glasgow" which was collected after 2005. This means two different mile 2 locations were sampled from 2006 to 2015. Samples were not collected at Broomielaw (mile 0) past December of 2005 which coincides with the 2006 completion of the Clyde Arc Bridge just west of the Broomielaw sampling point 1. Table 2.1 shows the frequency at which samples were taken at each location.

Samples from several depths were collected from each location. The sampling depths are for the most part constant over substantial time intervals but have changed a few times over the course of the river run. An example would be samples taken at 0 meters depth or surface samples. Surface water samples were taken from the beginning of 1970 to mid 2006. Subsequently the shallowest samples were taken at a depth of 1 meter. Table 2.2 shows the different depths and their frequencies.

Mile	Frequency	Mile	Frequency
0	2332	14	2933
2 Red	297	16	2929
2 Blue	2957	18	2875
4	3017	20	2463
6	3024	22	1493
8	3102	24 Red	2034
10	3013	24 Blue	42
12	2947	*	*

TABLE 2.1: The table above shows the frequency of the sample taken at each location. These frequencies reflect sampling across all 669 survey dates and all depths.

Depth (meters)	Frequency	Depth (meters)	Frequency
0	6541	5.5	9
1	1189	6	6412
1.35	1	6.5	2
1.5	7	7	35
2	6706	7.5	6264
2.3	1	8	51
2.5	21	8.5	1
3	6630	9	635
3.1	1	9.5	10
3.5	10	10	31
4	834	11	15
4.5	9	11.5	1
5	42	*	*

TABLE 2.2: The table above shows the frequency of the sample taken at each depth. These frequencies reflect sampling across all 669 survey dates and all locations.

The River Run data contains numerous variables and thus allows one to generate complex statistical models. Not all variables will be used in this research. However, mentioning a comprehensive list of variables may benefit any future research. These variables are shown in Table 2.3. While most of the variables in the River Run data are self explanatory, some variables warrant further description. These are as follows:

- Percent Through Tide Cycle (M2) -a value from a linear scale of the interval [0,100], where 0 corresponds to max/min tide and 100 corresponds to the subsequent min/max tide. To clarify, for an ebb tide state, the tide fraction will run from 0 at the start of the ebb tide to 100 at the end of the ebb tide. The subsequent flood tide state will see the tide fraction run from 0 at the start of the flood tide to 100 at the end of the flood tide.

- Days to Spring Tide - a value which runs from -8 to 8 where 0 represents spring tide, negative values represents days after spring tide, and positive values represent days before spring tide.

Data Variable	Units of Measurement (Factor Assignment *)
Dissolved Oxygen	milligrams per liter
Temperature	degrees Celsius
Salinity	parts per thousand ppt
Date of Sample	time and date
Tide State (Lunar Semi-Dihernal or M2)	ebb or flood *
Percent Through Tide Cycle (M2)	percentage
Days to Spring Tide	days
River Flow	cubic meters per second
Depth	meters
Distance From Tidal Wier	miles
Easting	British National Grid (not used)
Northing	British National Grid (not used)
Previous Tide Extreme (M2)	time and date (not used)
Next Tide Extreme (M2)	time and date (not used)

TABLE 2.3: The table above shows the variables collected during the River Run.

- River Flow - cubic meters per second(m^3/sec). The values correspond to the mean of the Daily Mean Flow from Daldowie gauging station of the date of the sample and the previous 4 days. Daily mean flows are calculated by taking measurements at 15 minute intervals from 9 am of the record date to 9 am of the following day.

There is a need to create new variables from the raw data for the purpose of extracting as much information as possible and to transform certain raw variables into variables conducive to statistical model construction. These are as follows:

- Day of Year (doy) - a whole number value from the cyclic interval [1,365].
- Year - a value with a fractional component to constitute a continuous time scale.
- Tide (M2) - adjusted from percent through tide cycle, tide is a value from the cyclic interval [0,200), where [0,100) represents low tide to high tide and [100,200) represents high tide to low tide. 0 coincides with 200.
- Spring - a whole number value in the cycle interval [0,16] where 0 and 16 represent neap tide and 8 represents spring tide.

The River Run sampling was not carried out in a completely random fashion. Factors such as technical difficulties, weather conditions, standardized hours of employment of SEPA personnel, and favorable tide conditions for boat travel may cause a certain level of bias in the data. It has also been mentioned by SEPA personnel that winter month sampling had been discontinued for several years due to the rationale that low dissolved oxygen levels are rare in colder weather and thus there is no need for monitoring. Figures 2.2-2.5 depict the sampling distribution by year, month, percent through tide cycle, and

days to spring tide respectively. Figure 2.6 shows that the time gap between sampling dates is also irregular.

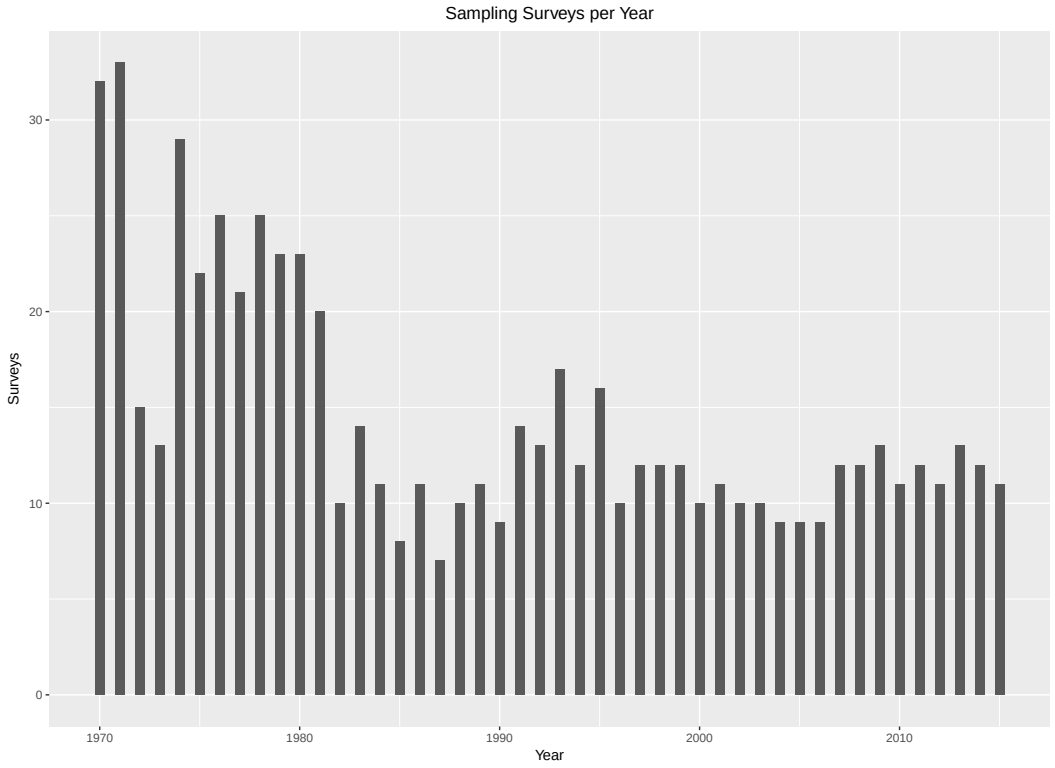


FIGURE 2.2: The number of River Run surveys for all years in the data set.

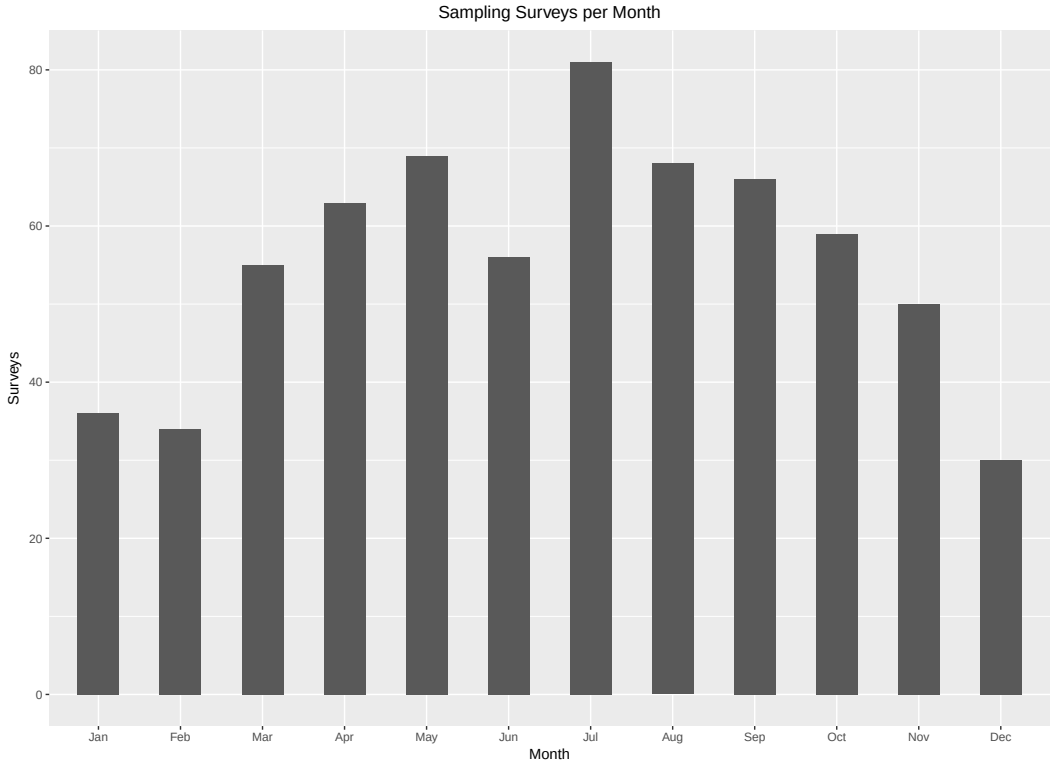


FIGURE 2.3: The number of River Run surveys for all months in the data set.

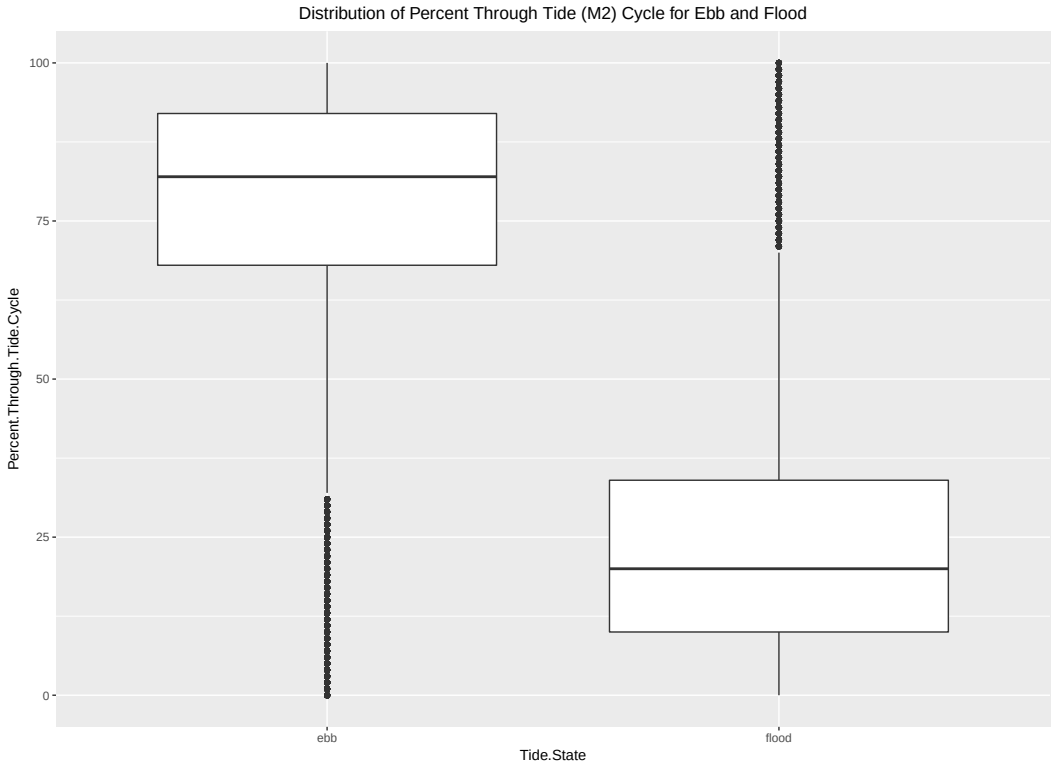


FIGURE 2.4: The distribution of Percent through Tide Cycle for both ebb and flood states in the data set. The median for ebb is 82 percent and the median for flood is 20 percent.

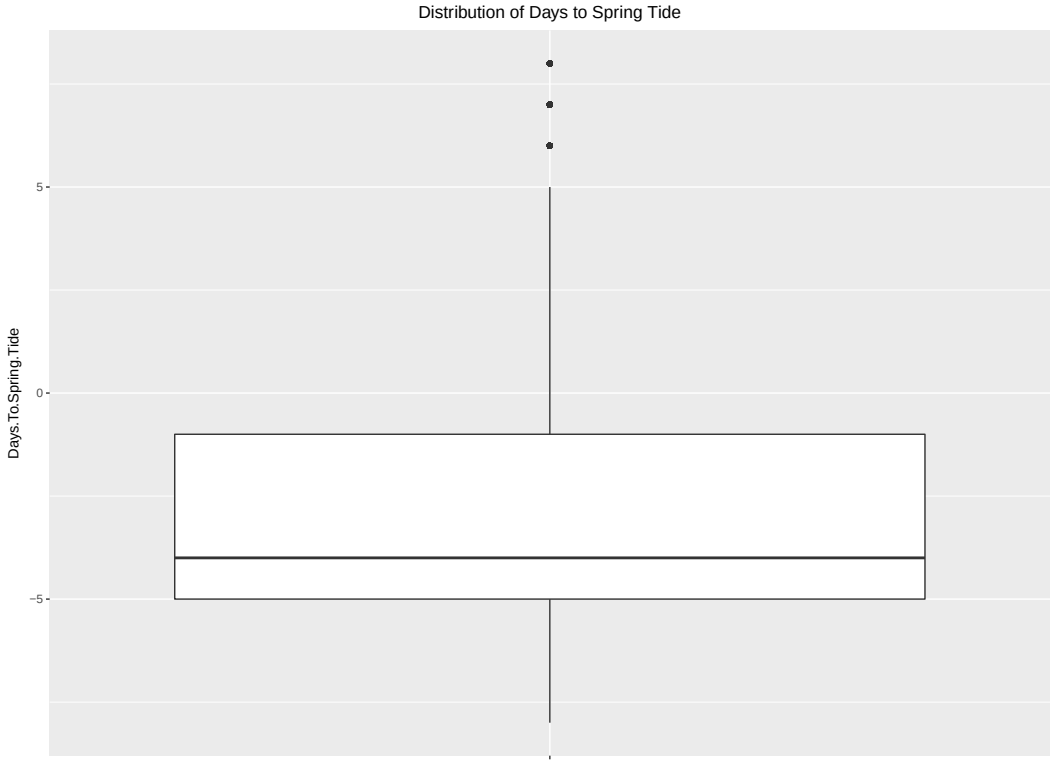


FIGURE 2.5: The distribution of Days to Spring Tide in the data set. The median is -4.

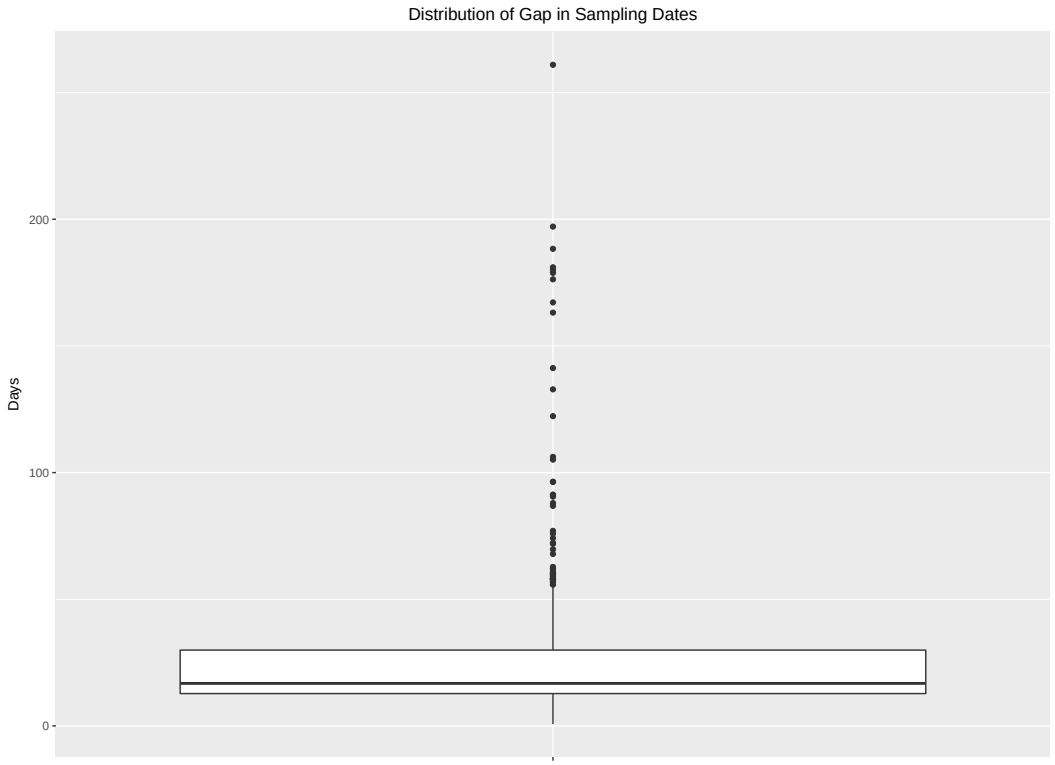


FIGURE 2.6: The distribution of the gap of sampling dates of the data set. The median is 16.8 days.

There are multiple issues with the River Run data set. These are as follows:

- The multiple locations for Mile 2 and Mile 24 may cause an elevated variance.
- The inclusion of depths less than 2 meters from the surface in order to have a data set which spans all years, may cause an elevated variance.
- More sampling took place in the early years.
- More sampling took place in the warmer months.
- Most sampling took place in the late stages of the ebb state and the early stages of the flood state.
- Most sampling took place after spring tide.

Nonetheless, the sample size is large and a suitable model should be able to overcome many of the non-uniform patterns in the covariates.

2.3 Relationships Between Potential Explanatory Covariates and Dissolved Oxygen

The River Run data is a rich data set which continues to be augmented to this day. Although river samples are taken at multiple depths, this analysis will be restricted to depths less than 2 meters. The ground work being laid here can be extended to other depths.

Table 2.3 lists the raw data variables, some of which need to be transformed slightly in order to create more meaningful values in the context of a statistical model. Table 2.4 depicts the covariates which will be used in the River Run model.

Scatterplots of each explanatory covariate against the dissolved oxygen give us an initial sense of the underlying relationships. These scatterplots are depicted in Figure 2.7. Some of these relationships are not surprising, while others may not be as intuitive. These are summarized below.

- High temperatures tend to cause the dissolved oxygen to drop in fresh water Ringler and Hall (1975). This is evident in panels a) and d) of Figure 2.7.
- The solubility of oxygen decreases as salinity increases Kielmas (2019). Panel e) of Figure 2.7 shows this to be the case for the salinity between 0 and approximately 15 ppt. The gradual increase in dissolved oxygen for salinity values above 15 ppt

Variables	Units of Measurement or Description
Dissolved Oxygen	milligrams per liter
Day of Year (doy)	a whole number value contained in [1,365] where 1 represents 1 January and 365 represents 31 December
Station	miles from the tidal weir
Year	a value with a fractional component to constitute a continuous time scale
Temperature	degrees Celsius
Salinity	parts per thousand
Tide	a whole number value contained in [0,200] where 0 represents the start of the flood tide, 100 represents the end of the flood tide along with the start of the ebb tide, and 200 represents the end of the ebb tide
Spring	a whole number value contained in [0,16] where 0 and 16 represent neap tide and 8 represents spring tide
River Flow	cubic meters per second
Survey	a factor denoting the sampling date to be used as a random effect

TABLE 2.4: Table of the variables to be used for the River Run model.

can be attributed to brackish water samples taken well downstream from the city centre and in close proximity to the Firth of Clyde. These samples being closer to open water would inherently have higher levels of dissolved oxygen. Panel b) of Figure 2.7 also exhibits this pattern.

- Panel h) of Figure 2.7 exhibits an increase in dissolved oxygen as river flow increases. It seems reasonable that substantial rainfall and subsequent increases in river flow would perturb the water and introduce more dissolved oxygen.
- We also clearly see a general increase in dissolved oxygen as the years go by in panel c) of Figure 2.7.
- Plots f) and g) of Figure 2.7 depicting Tide and Spring are the only two which seem to not have any clear discernible patterns.

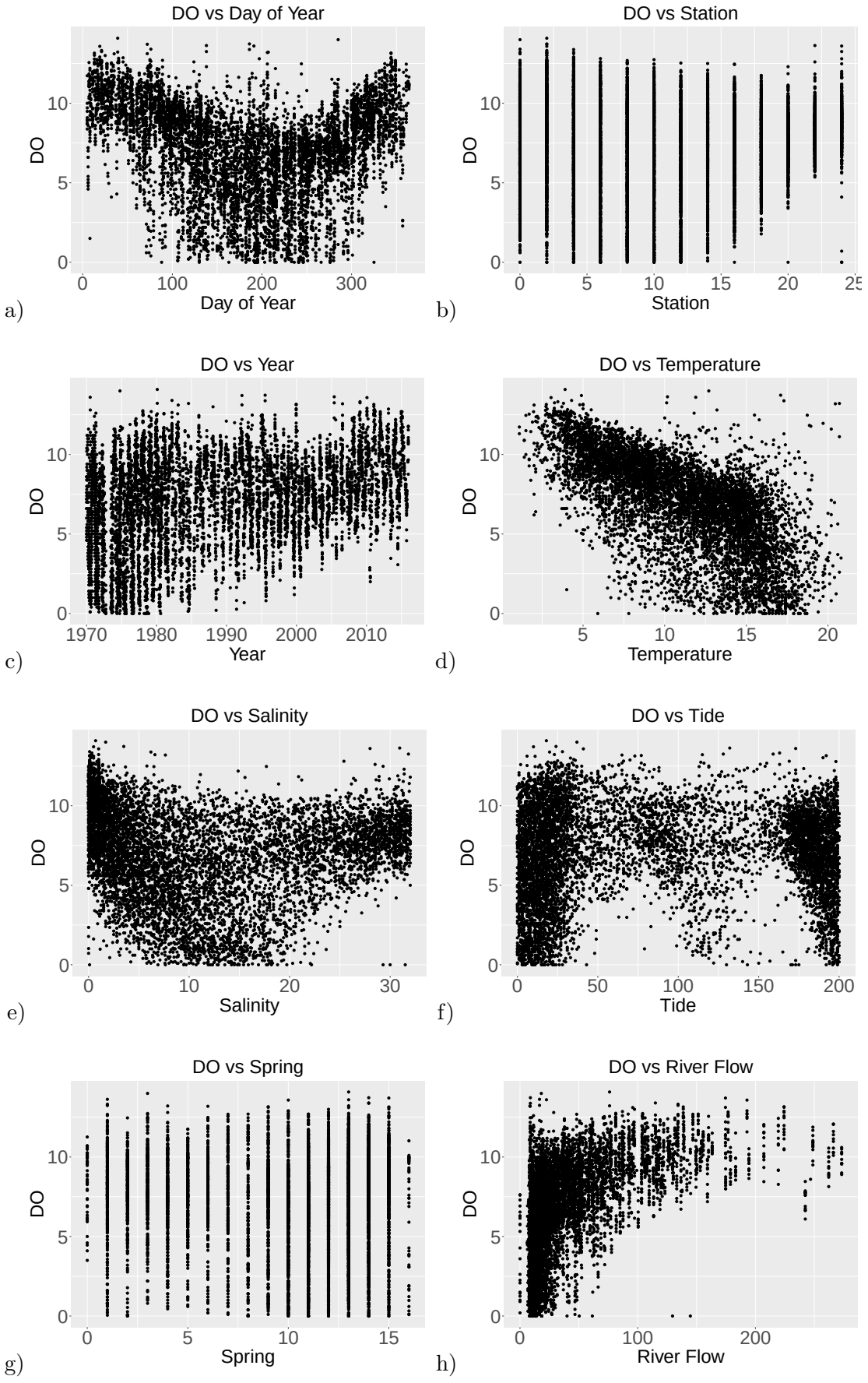


FIGURE 2.7: Plots of DO vs explanatory covariates.

2.4 Relationships Between Potential Explanatory Covariates and Other Potential Explanatory Covariates

The relationships of each explanatory covariate with the dissolved oxygen are not the only relationships of interest. Explanatory covariates plotted against each other provide insights to the relationnships between covariates which may be present. Figures 2.8–2.10 depict the pairwise relationships of most interest.

- Panel a) of Figure 2.8 shows patterns of diagonal sets of points consistent with the sampling throughout Year taking place at different values of Day of Year. There also seems to be fewer samples taken in the colder months between 2000 and 2005.
- Panel b) of Figure 2.8 shows sampling was halted at Station 0 in 2006 and at Stations 20, 22, and 24 in 2010.
- Panel c) of Figure 2.8 shows the periodic nature of Temperature through the seasons.
- Panel d) of Figure 2.8 shows a generally uniform level of Salinity through Years with the exception of a bit more values of lower salinity.
- Panel e) of Figure 2.8 shows most samples from 1972 to 2000 were taken during low diurnal Tide. Sampling was more evenly distributed throughout the diurnal Tide cycle after 2000.
- Panel f) of Figure 2.8 shows sampling was relatively uniformly distributed from 1970 to 1980 and from 2000 to 2016. Alternatively, sampling took place mostly 2 days after Spring tide approaching neap tide.
- Panel g) of Figure 2.8 shows mostly low River Flow with some higher River Flow scattered throughout the years.
- Panel h) of Figure 2.8 shows Temperature ranges decrease as samples are taken further down stream.
- Panel a) of Figure 2.9 shows that Salinity increases as samples are taken further down stream.
- Panel b) of Figure 2.9 shows that a majority of samples were taken at low Tide.
- Panel c) of Figure 2.9 shows that Temperature is higher in the middle of the year.
- Panel d) of Figure 2.9 shows that more samples were taken in the middle of the year.

- Panel e) of Figure 2.9 shows most samples were taken at low Tide with some taken at high Tide during the middle of the year.
- Panel f) of Figure 2.9 shows a majority of samples were taken as spring tide turned into neap tide.
- Panel g) of Figure 2.9 shows there was higher River Flow at the beginning and end of the year.
- Panel h) of Figure 2.9 shows no discernible relationship between Spring and Temperature other than slight periodic behavior.
- Panel a) of Figure 2.10 shows there is a gap in sampling at high diurnal Tide as Spring moves from spring tide to neap tide.
- Panel b) of Figure 2.10 shows a few values of higher River Flow as Spring moves from spring tide to neap tide.
- Panel c) of Figure 2.10 shows most sampling took place during low Tide.
- Panel d) of Figure 2.10 shows most was started at low Tide with high Salinity and as the Tide started rising the Salinity dropped. This may have been caused by the sampling boat starting at the Station furthest from the city center and traveled upstream as it took samples.
- Panel e) of Figure 2.10 shows that Temperature tends to drop as the River Flow increases.
- Panel f) of Figure 2.10 shows that Salinity tends to drop as the River Flow increases.
- Panel g) of Figure 2.10 shows that Temperature ranges decreases as the Salinity increases.

One important feature of the data is the periodic nature of some of the explanatory variables. Day of Year, Tide, and Spring all rise and fall with each having a different period. Furthermore, all three are nested within the Year explanatory variable; Tide is contained in Spring, which is contained in Day of Year, which is contained in Year.

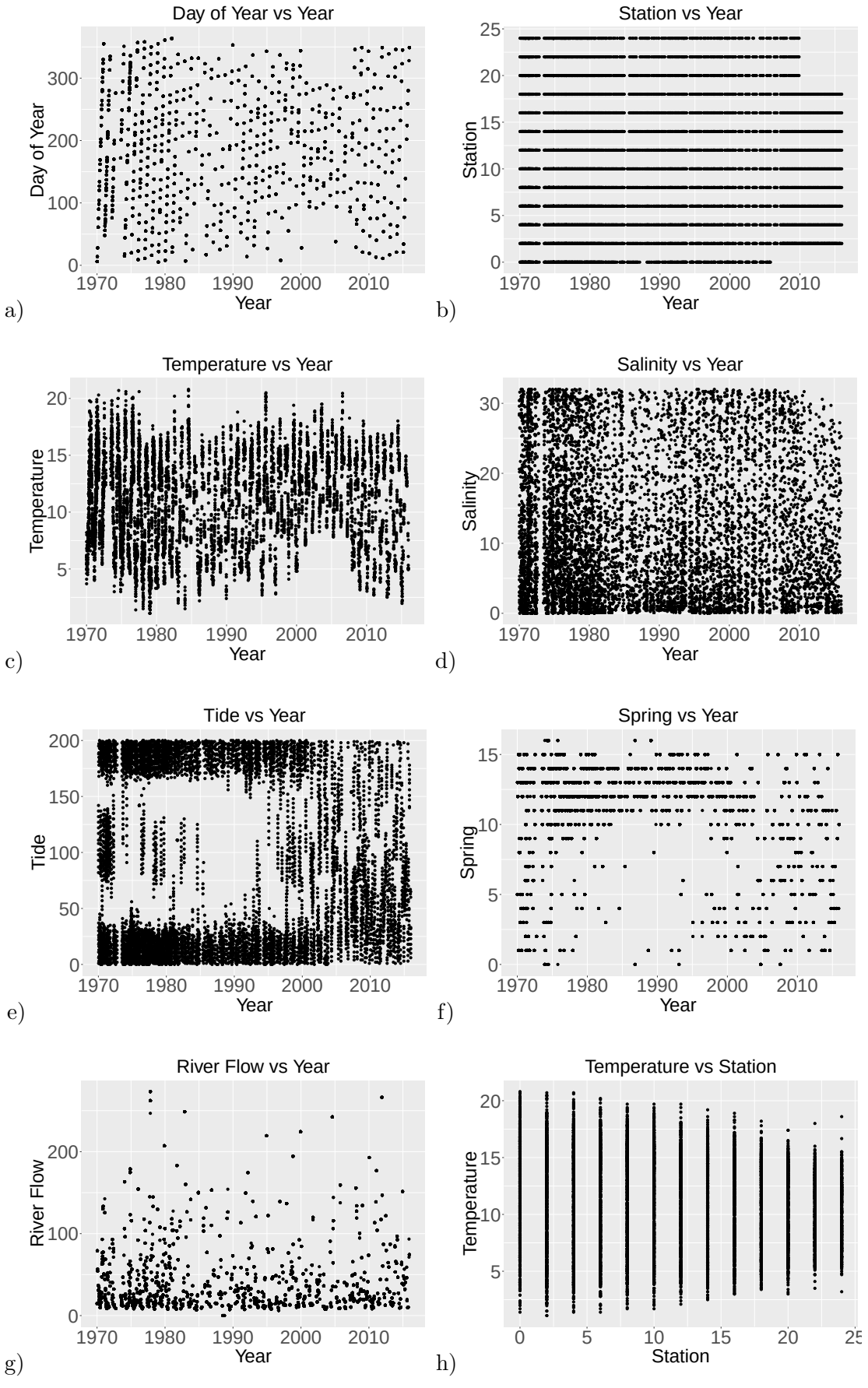


FIGURE 2.8: Plots of explanatory covariates vs other explanatory covariates.

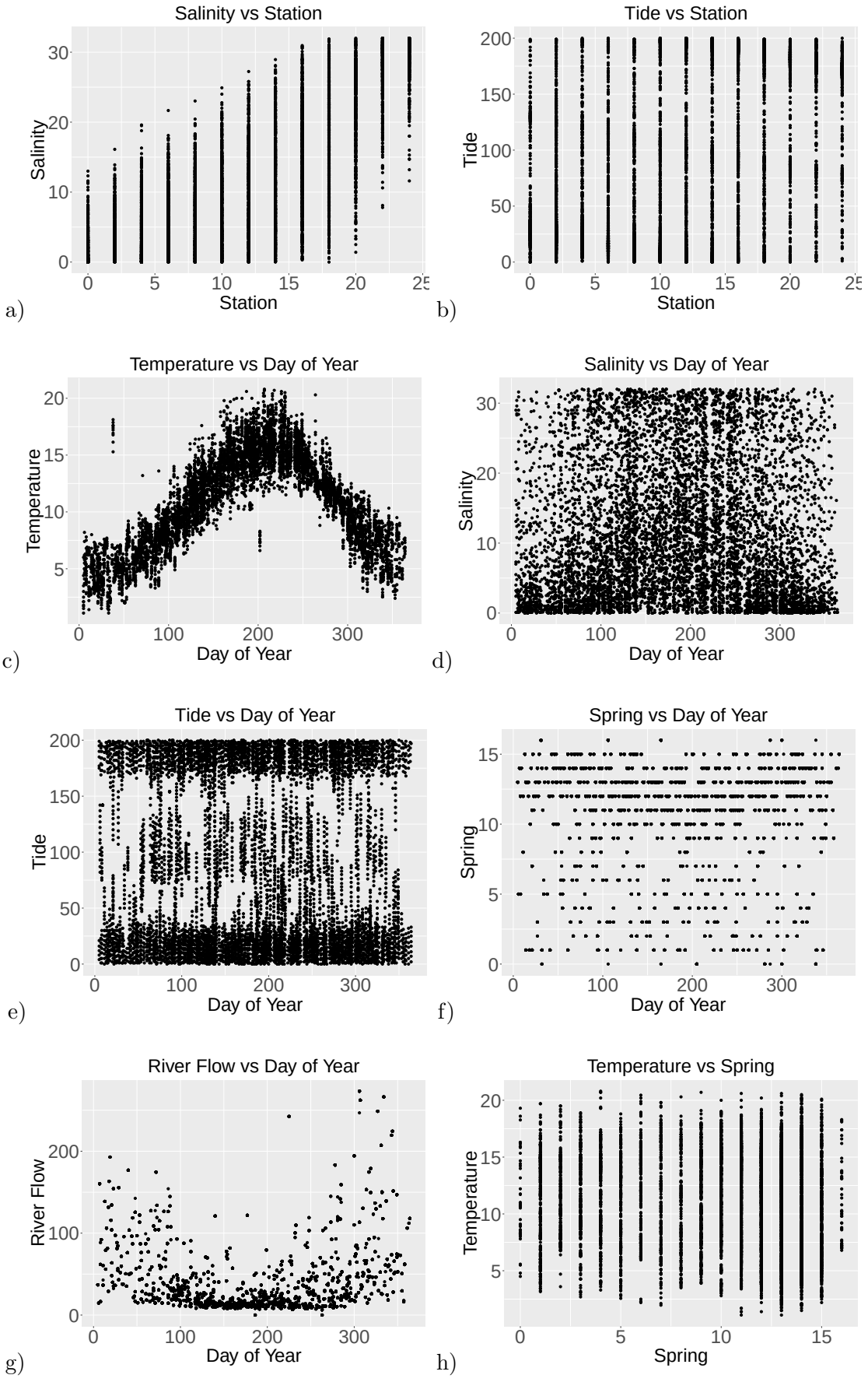


FIGURE 2.9: Plots of explanatory covariates vs other explanatory covariates.

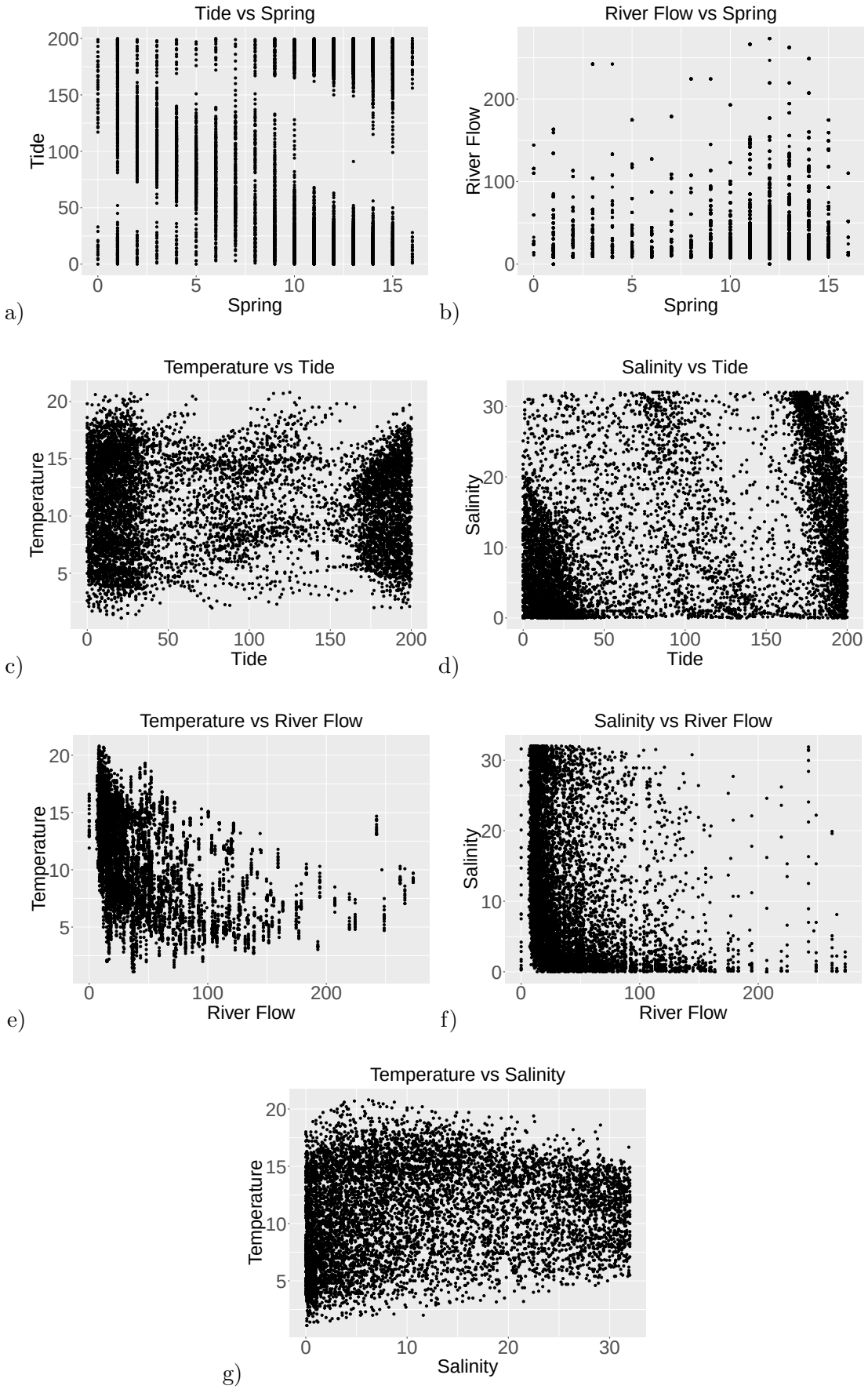


FIGURE 2.10: Plots of explanatory covariates vs other explanatory covariates.

2.5 Time Series Plots for Individual Stations

The dissolved oxygen in the River Clyde has been steadily increasing since 1970. This is evident upon review of panel c) of Figure 2.7. However, all stations are included in this plot. Looking at the dissolved oxygen over the years at each individual location along the river gives a clearer picture of the improvement in water quality. Stations close to the city centre, where more human activity may have contributed to the depletion of the dissolved oxygen, may have seen a more dramatic increase compared to stations further downstream.

Time series plots of the dissolved oxygen values for all Stations are depicted in Figures 2.11-2.12. Superimposed on these plots are smooth functions of Year, smooth functions of Year and Day of Year, and smooth functions of Year and Day of Year with an interaction term included. It appears the most dramatic increases in oxygen are happening at Stations 4 through 12. Furthermore, Stations 22 and 24 are seeing relatively little increase over the years.

The smooth curves also tell an interesting story. There seems to be a pronounced adjustment from the smooth of Year and Day of Year without the interaction term (shown in red) and the smooth of Year and Day of Year with the interaction term (shown in blue) for the higher values of Year for most Stations. Station 2 depicted in panel b) of Figure 2.11 is a good example where the interaction terms makes a minor adjustment, whereas Station 10 depicted in panel f) of Figure 2.11 is a good example where the interaction terms makes a drastic adjustment. This can be better seen in Figure 2.13. These contour plots depict the adjustments made by the interaction term to the marginal effects by means of the contours along with the overall oxygen levels by means of the colours in the background. The detailed construction of these contour plots will be described later in this thesis. The lack of contours for Station 2 in panel a) of Figure 2.13 is consistent with the tendency of the two periodic smooth functions in panel b) of Figure 2.11 not straying too far from each other. Furthermore, the abundance of contours for Station 10 in panel b) of Figure 2.13 is consistent with the tendency of the two periodic smooth functions in panel f) of Figure 2.11 straying a good deal from each other at the low and high end of Year. This is compelling evidence these interactions with Year and other explanatory covariates will be warranted when constructing a representative additive model. Furthermore, the different sizes of adjustments across Station implies the need for interaction term involving Station and other explanatory covariates.

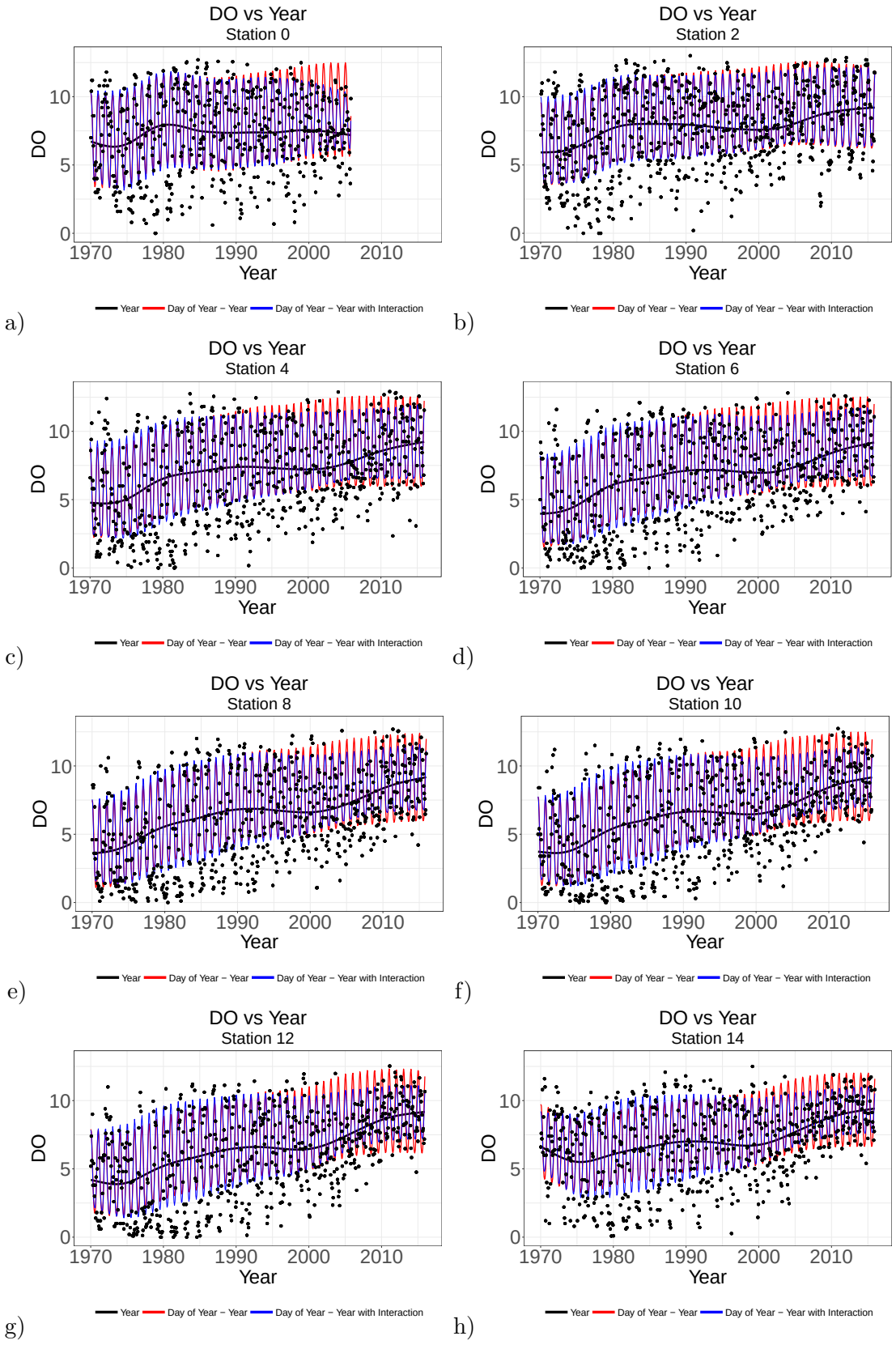


FIGURE 2.11: Time series plots for Stations 0 through 14 with Year and Day of Year smooths.

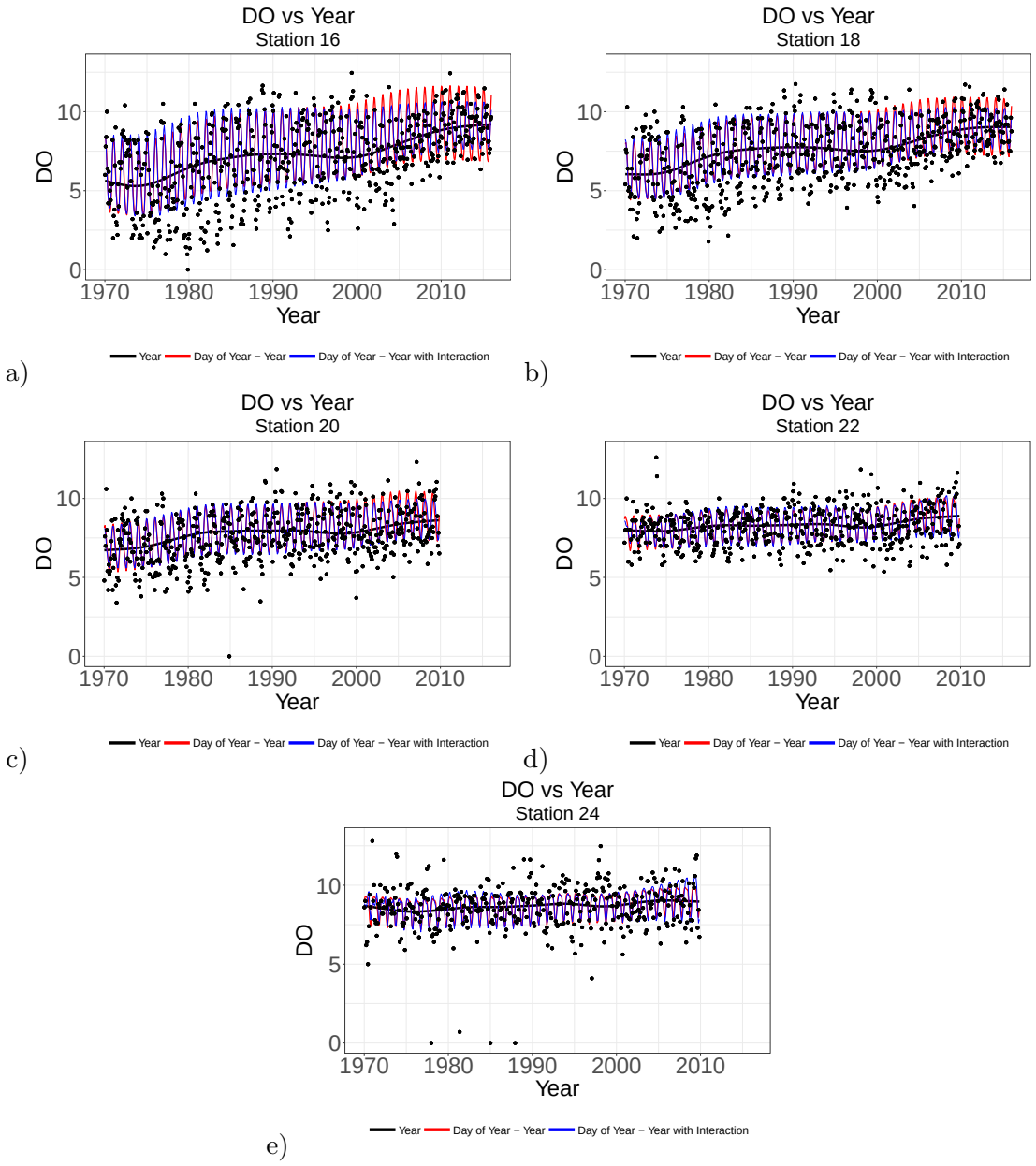
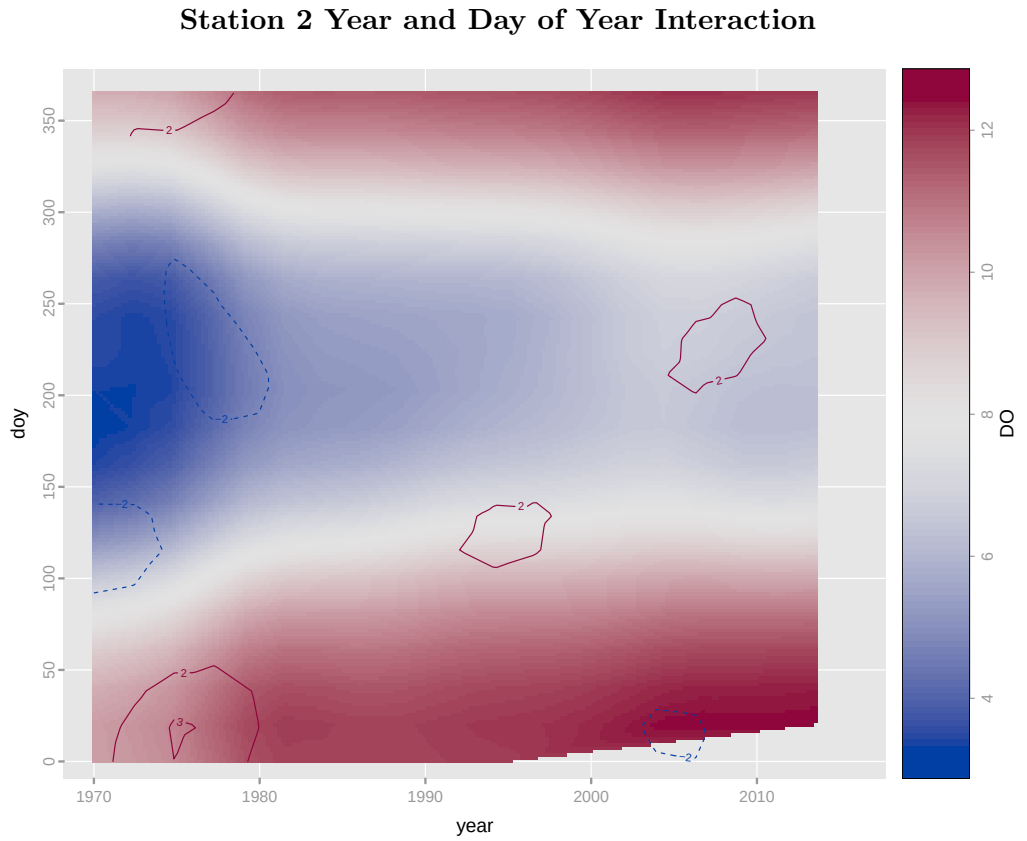
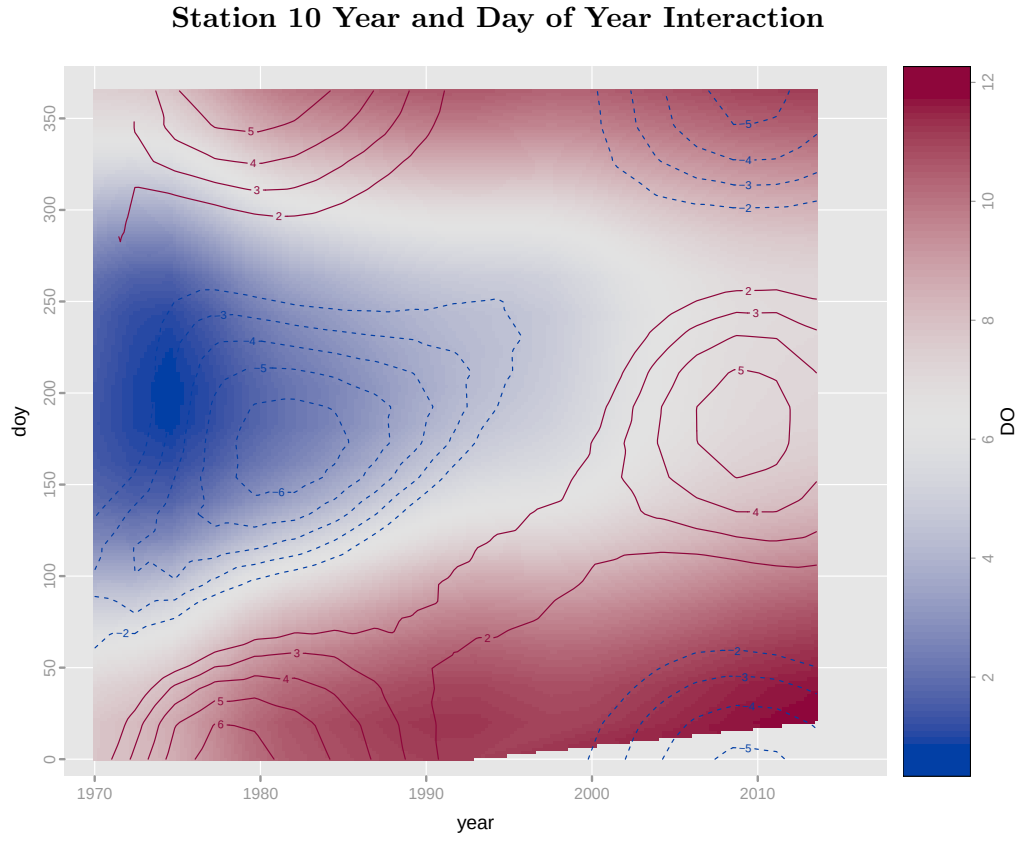


FIGURE 2.12: Time series plots for Stations 16 through 24 with Year and Day of Year smooths .



a)



b)

FIGURE 2.13: Contour plot of interaction term for smooth function of Year and Day of Year with interaction for Station 10.

2.6 Summary

The River Run data set is complex, with some issues to be addressed. Not all values of the explanatory variables are represented uniformly and consistently. Some months, years, and tide stages are represented more than others. There is however a large amount of data which contains valuable information. The explanatory covariates singled out show distinct relationships with the dissolved oxygen. There are also notable relationships amongst the explanatory covariates themselves. The time series plots representing each station provided insight showing the level of water quality improvement along the river. Plots of multiple Stations, along with the inclusion of various smooth functions, support the inclusion of interaction terms in an additive mixed model. This will be addressed in Chapter 5. However, before an additive mixed model can be constructed, the methodology for fitting said model must first be discussed.

Chapter 3

Additive Models and Interactions

The number of data fields, the sampling regimes, and the sizes of the River Run data set provide motivation to generate complex yet representative models of the dissolved oxygen in the River Clyde. Parametric models fail more times than not in capturing the complexity of environmental data sets such as these. Additive models, which include smooth functions of some explanatory covariates, offer the necessary flexibility required in this instance. Furthermore the sampling regimes of the River Run data set warrant the implementation of a random effects component. Chapter 3 offers some background for these methodologies. There is particular emphasis on interactions because of the complexity of the models which will be required.

One would expect the dissolved oxygen would not change substantially for small increments of the explanatory covariates. Furthermore, the underlying drivers of the dissolved oxygen most likely also influence other biological and chemical processes which in turn affect the levels of oxygen. Since the vast majority of these processes are not accounted for in the data sets, the effect of the available covariates may be better represented by flexible functions. Smooth functions allow for these small changes and provide the flexibility needed to capture the underlying complex relationships. There are several methods of smoothing outlined in the literature. Some examples taken from Wood (2017) are cubic regression splines, cyclic cubic regression splines, basis splines (B-splines), and penalized basis splines (P-splines). The type of smoothing is not of primary importance, so there is the benefit to choosing the method which is most convenient. Eilers and Marx (1996) introduce P-splines, which penalize the second differences of the B-spline coefficients, and thus are easily implemented. In order to fully describe P-splines there must be detail given regarding B-splines.

3.1 *B*-Splines

Introduced by Schoenberg (1946), *B*-splines are well conditioned (De Boor (1972)), and can be described as a collection of polynomial functions which are joined together in a special way (Eilers and Marx (1996)). These polynomial pieces are joined at points called *knots* along the independent axis. The knots can be equally spaced, or spaced according to quantiles, or can use some other spacing criterion. At this point it is worth noting subsequent work by Eilers and Marx (2010) which compared the two advocated approaches to using *B*-splines: 1) the use of a *B*-spline basis, equally spaced knots, and difference penalties, and 2) the use of truncated power functions, knots based on quantiles, and a ridge penalty. The authors found approach 1) to be clearly preferred and thus will be the method implemented in this thesis.

3.1.1 1-Dimensional Case

The customary notation representing a single *B*-spline of polynomial degree q evaluated at independent covariate value x is

$$B(x; q), \tag{3.1}$$

or more specifically as one component of a basis composed of many *B*-splines,

$$B_j(x; q) \tag{3.2}$$

for x values lying between knots j and $j + 1$. As mentioned by Eilers and Marx (1996), the general properties of a *B*-spline of degree q are:

- it consists of q+1 polynomials of degree q
- the polynomial pieces are joined by q inner knots
- at the joining points, derivatives of up to order q-1 are continuous
- the *B*-spline is positive on a domain spanned by q+2 knots and zero everywhere else.

Figure 3.1 depicts single *B*-splines of order 1, 2, 3, and 4, all of which span the interval [0,1]. Single *B*-splines of degree q contain q+2 evenly spaced knots. It is plain to see that all of the above properties are satisfied.

Now that a single B-spline has been constructed, the next step is to construct a basis composed of numerous B-splines. Wood (2017) defines a B-spline basis composed of k individual B-splines, or a k parameter B-spline basis, as follows:

- Define $k + q + 1$ knots $x_1 < x_2 < ... < x_{k+q+1}$.
- The interval over which the B-splines are to be evaluated lie within $[x_{q+1}, x_{k+1}]$, thus making the first and last q knots arbitrary.
- A q order B-spline can then be represented as

$$f(x) = \sum_{j=1}^{k} B_j(x; q)\alpha_j \tag{3.3}$$

  where the B-spline basis functions are defined recursively as

$$B_j(x; q) = \frac{x - x_j}{x_{j+q+1} - x_j} B_j(x; q-1) + \frac{x_{j+q+2} - x}{x_{j+q+2} - x_{j+1}} B_{j+1}(x; q-1) \tag{3.4}$$

  for j=1,...k and

$$B_j(x; 0) = \begin{cases} 1 & x_j \leq x < x_{j+1} \\ 0 & \text{otherwise} \end{cases} \tag{3.5}$$

Also mentioned by Eilers and Marx (1996), general properties of a B-spline basis are

- the B-spline overlaps with $2q$ polynomial pieces of its neighbours, except at the boundaries
- at a given x, q+1 B-splines are non-zero.

The second of the above mentioned properties makes B-splines strictly local, which is appealing (Wood (2017)). This prevents individual points from having global effects. Figure 3.2 depicts several B-spline bases of orders 1 through 4. The points depict the evenly spaced knots. Once again it is plain to see the properties mentioned above are present. Furthermore, when $\alpha_1 = \alpha_2 = ... = \alpha_k = 1$, the B-spline values at any x within the interval $[x_{q+1}, x_{k+1}]$ sum to 1.

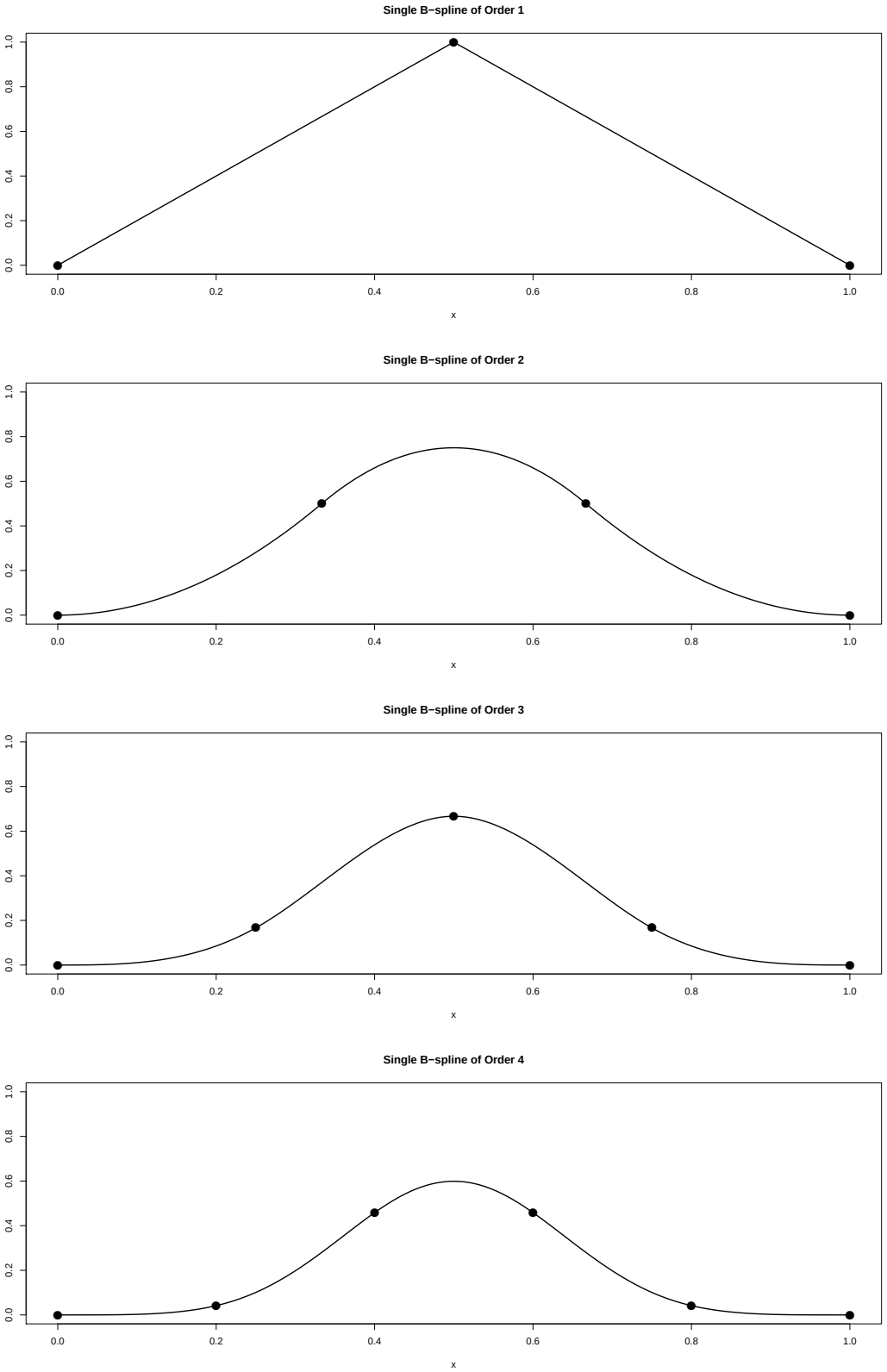


FIGURE 3.1: The above plots show single B-splines of order 1 through 4. The points represent the knots where the polynomial pieces are joined. It is evident all properties of single B-splines mentioned previously are present.

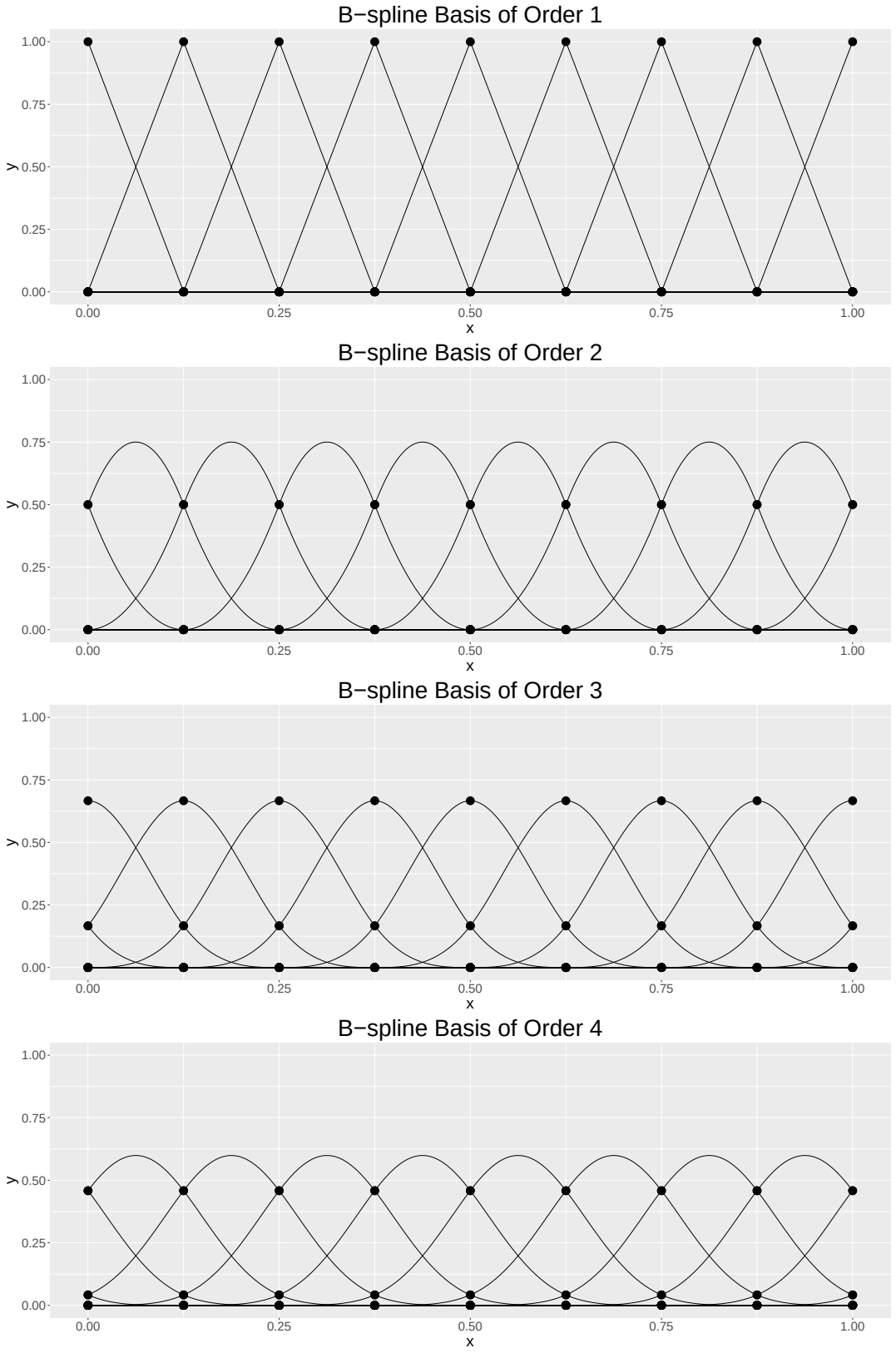


FIGURE 3.2: The above plots show B-spline bases of order 1 through 4. The points represent the knots where the polynomial pieces are joined.

3.1.2 2 and 3-Dimensional Case

The obvious next step is to apply these B-spline principles to a smooth function of two explanatory covariates. This can be done naturally by using tensor products (Eilers et al. (2006)). Consider the ordered triplets (x_{1_i}, x_{2_i}, y_i) for $i = 1, ..., n$. The tensor product of a single B-spline in x_1, say the j^{th}, and a single B-spline in x_2, say the l^{th}, can be easily represented as follows:

$$B_{j,l}(x_1, x_2; q) = B_j(x_1; q)B_l(x_2; q). \tag{3.6}$$

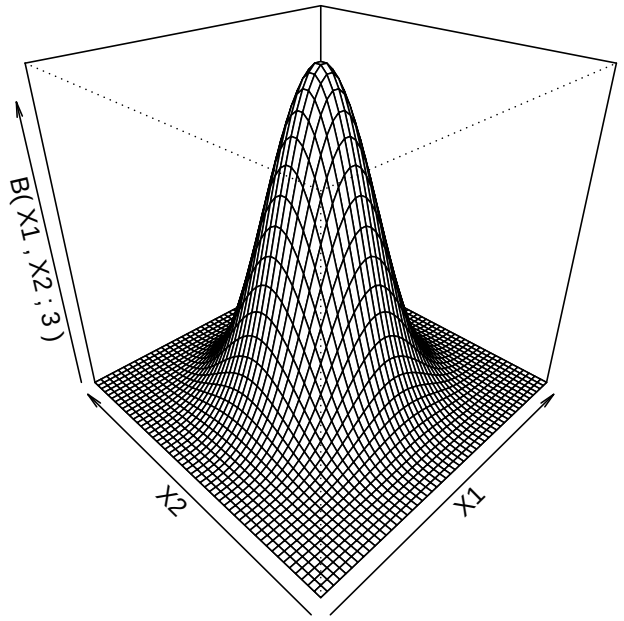


FIGURE 3.3: The above plot shows a 2-dimensional B-spline of order 3 as the tensor product of two 1-dimensional B-splines.

The result is a surface as depicted in Figure 3.3. In this case third order B-splines are being used ($q = 3$). Figure 3.4 depicts a 2-dimensional B-spline basis over a grid composed of x_1 and x_2 is generated when j and l are indexed $j = 1, ..., k_1$ and $l = 1, ..., k_2$, where k_1 and k_2 correspond to the number of parameters (basis functions) associated with x_1 and x_2 respectively.

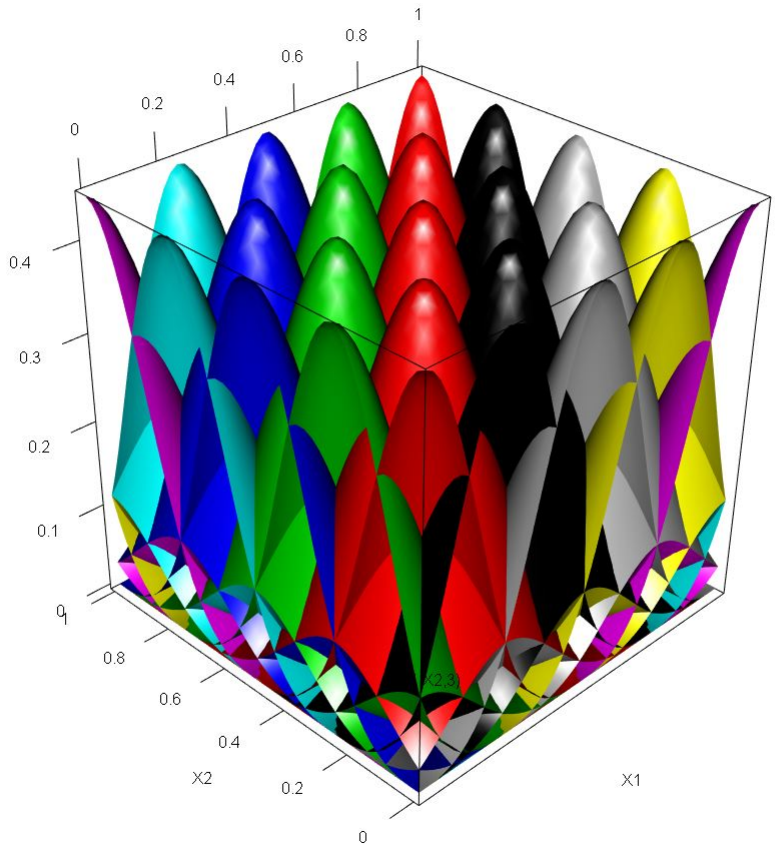


FIGURE 3.4: A 2-dimensional B-spline basis of order 3 with $k_1 = k_2 = 5$

In the 1-dimensional case, the coefficients or parameters represented by the α's are elements of a vector. For the 2-dimensional case, the coefficients can be represented as elements of a matrix (Eilers et al. (2006)). Let $A = [\alpha_{jl}]$ be a $k_1 \times k_2$ matrix of such coefficients. A B-spline function of two variables can now be expressed as:

$$\begin{aligned} f(x_1, x_2) &= \sum_{j=1}^{k_1} \sum_{l=1}^{k_2} B_{jl}(x_1, x_2; q)\alpha_{jl} \\ &= \sum_{j=1}^{k_1} \sum_{l=1}^{k_2} B_j(x_1; q)B_l(x_2; q)\alpha_{jl}. \end{aligned} \tag{3.7}$$

where the α_{jl}'s scale the individual 2-dimensional B-splines.

The same concept can now be extended to a smooth B-spline function of three variables. Consider the ordered quadruples $(x_{1_i}, x_{2_i}, x_{3_i}, y_i)$ for $i = 1, ..., n$. The tensor product of a single B-spline in x_1, say the j^{th}, a single B-spline in x_2, say the l^{th}, and a single

B-spline in x_3, say the m^{th} can be, similarly to equation 3.6, represented as follows:

$$B_{j,l,m}(x_1, x_2, x_3; q) = B_j(x_1; q)B_l(x_2; q)B_m(x_3; q) \tag{3.8}$$

where j, l, and m are indexed $j = 1, ..., k_1$, $l = 1, ..., k_2$, and $m = 1, ..., k_3$ where k_1, k_2, and k_3 correspond to the number of parameters (basis functions) associated with x_1, x_2, and x_3 respectively.

The 1-dimensional and 2-dimensional B-spline functions can have the α's represented as elements of a vector and a matrix respectively. B-spline functions of three variables will naturally contain coefficients which can be represented as elements of a 3-dimensional array. Let $A = [\alpha_{jlm}]$ be a $k_1 \times k_2 \times k_3$ array of such coefficients. A B-spline function of three variables can now be expressed as:

$$\begin{aligned} f(x_1, x_2, x_3) &= \sum_{j=1}^{k_1} \sum_{l=1}^{k_2} \sum_{m=1}^{k_3} B_{jlm}(x_1, x_2, x_3; q)\alpha_{jlm} \\ &= \sum_{j=1}^{k_1} \sum_{l=1}^{k_2} \sum_{m=1}^{k_3} B_j(x_1; q)B_l(x_2; q)B_m(x_3; q)\alpha_{jlm}. \end{aligned} \tag{3.9}$$

where the α_{jlm}'s scale the individual 3-dimensional B-splines. In the interest of notational brevity, $B_j(x_j)$ will be used instead of $B_j(x_j, q)$ when B-spline order q is implied or contextually unimportant.

3.2 *P*-Splines

P-splines are a means of reducing model complexity by imposing a penalty on B-spline coefficients. The following section introduces P-splines by initially considering univariate functions and expanding to bivariate and trivariate functions.

3.2.1 1-Dimensional Case

Different types of smoothing involve the inclusion of a smoothness penalty (Eilers et al. (2015)). This penalty is generally composed of the integrated square of the first or higher orders of derivative of the function, but commonly penalizes the function's curvature by using the integrated square of the function's second derivative, although there is nothing special about the second derivative (Eilers and Marx (1996)). The customary

least squares term

$$\sum_{i=1}^{n}[y_i - f(x_i)]^2 \tag{3.10}$$

will have added to it this penalty to obtain the objective function to be minimized. That is to say the smooth function is generated by minimizing the objective function

$$S = \sum_{i=1}^{n}[y_i - f(x_i)]^2 + \lambda \int [f^{(d)}(x)]^2 dx \tag{3.11}$$

for some non-negative smoothing parameter λ and derivative order d. The parameter λ controls the amount of smoothing. For values of λ close to zero, f is very erratic or "wiggly" producing a high variance. As λ approaches infinity, f will become linear, producing a high bias. The selection of the smoothing parameter for an acceptable bias-variance trade off most commonly uses ordinary cross-validation or generalized cross-validation. Controlling smoothing will be discussed in detail in Section 3.5.

Eilers and Marx (1996) introduce penalized B-splines, or P-Splines, where the second term in equation (3.11), known as the penalty term, is replaced by a penalty based on differences of order d applied to the α's.

The difference operator Δ is defined as follows:

$$\Delta z_j = \Delta^{(1)} z_j = z_j - z_{j-1} \tag{3.12}$$

for some list of elements $(z_1, z_2, ..., z_J)$ and for $j = 2, .., J$. This is known as the *first* difference. Higher order differences are calculated by repeatedly applying the first difference operator. For example, the second difference can be calculated as follows:

$$\begin{aligned} \Delta^{(2)} z_j &= \Delta(\Delta z_j) \\ &= (z_j - z_{j-1}) - (z_{j-1} - z_{j-2}) \\ &= z_j - 2z_{j-1} + z_{j-2} \end{aligned} \tag{3.13}$$

for $j = 3, ..., J$. The d order differencing can be carried out on a vector of values $\boldsymbol{z} = (z_1, ..., z_J)^T$ by the matrix multiplication $\boldsymbol{D_d z}$, where $\boldsymbol{D_d}$ is a $(J-d) \times J$ differencing

matrix. A first example of a differencing matrix is

$$\boldsymbol{D_1} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & -1 & 1 & 0 \\ 0 & \dots & 0 & -1 & 1 \end{bmatrix} \tag{3.14}$$

where $\boldsymbol{D_1}$ is a $(J-1) \times J$ matrix. A second example is

$$\boldsymbol{D_2} = \begin{bmatrix} 1 & -2 & 1 & 0 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & \dots & 1 & -2 & 1 & 0 \\ 0 & \dots & 0 & 1 & -2 & 1 \end{bmatrix} \tag{3.15}$$

where $\boldsymbol{D_2}$ is a $(J-2) \times J$ matrix (Eilers et al. (2006)).

The differencing now needs to be applied to the coefficients of the B-spline basis, the α's. Let $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k)^T$ be a vector of B-spline basis coefficients and $\boldsymbol{D_d}$ be the differencing matrix corresponding to the difference operator $\Delta^{(d)}$. Following Eilers and Marx (1996), the objective function to minimize becomes

$$S = \sum_{i=1}^{n} [y_i - f(x_i)]^2 + \lambda \sum_{j=d+1}^{k} (\Delta^{(d)} \alpha_j)^2. \tag{3.16}$$

For example, for $d = 1$, the penalty term in equation 3.16 would be

$$\lambda \sum_{j=2}^{k} (\alpha_j - \alpha_{j-1})^2 \tag{3.17}$$

For $\boldsymbol{x} = (x_1, ..., x_n)^T$, where n is the number of observations, the B-spline function f can be expressed as

$$f(\boldsymbol{x}) = \boldsymbol{B\alpha} \tag{3.18}$$

where $\boldsymbol{B}$ is an $n \times k$ B-Spline design matrix and $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_k)^T$ is the coefficient vector. Equation (3.16) can now be expressed in terms of $\boldsymbol{B}$, $\boldsymbol{\alpha}$, and difference matrix $\boldsymbol{D_d}$:

$$S = \|\boldsymbol{y - B\alpha}\|^2 + \lambda \|\boldsymbol{D_d \alpha}\|^2 \tag{3.19}$$

where $\boldsymbol{y} = (y_1, ..., y_n)^T$. Using the identity

$$\|\boldsymbol{z}\|^2 = \boldsymbol{z}^T\boldsymbol{z} \tag{3.20}$$

for some column vector $\boldsymbol{z}$, equation 3.19 can be written as

$$S = (\boldsymbol{y} - \boldsymbol{B\alpha})^T(\boldsymbol{y} - \boldsymbol{B\alpha}) + \lambda\boldsymbol{\alpha^T D_d^T D_d \alpha}. \tag{3.21}$$

Upon expanding and subsequent factoring, the right side of equation 3.21, the objective function becomes

$$\begin{aligned} S &= \boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{\alpha}^T\boldsymbol{B}^T\boldsymbol{y} + \boldsymbol{\alpha}^T\boldsymbol{B}^T\boldsymbol{B\alpha} + \lambda\boldsymbol{\alpha}^T\boldsymbol{D_d}^T\boldsymbol{D_d\alpha} \\ &= \boldsymbol{y}^T\boldsymbol{y} - 2\boldsymbol{\alpha}^T\boldsymbol{B}^T\boldsymbol{y} + \boldsymbol{\alpha}^T(\boldsymbol{B}^T\boldsymbol{B} + \lambda\boldsymbol{D_d}^T\boldsymbol{D_d})\boldsymbol{\alpha} \end{aligned} \tag{3.22}$$

Minimizing S requires taking it's first derivative with respect to $\boldsymbol{\alpha}$ and setting that first derivative equal to zero.

$$\begin{aligned} \frac{dS}{d\boldsymbol{\alpha}} = 0 &= -2\boldsymbol{B}^T\boldsymbol{y} + 2(\boldsymbol{B}^T\boldsymbol{B} + \lambda\boldsymbol{D_d}^T\boldsymbol{D_d})\hat{\boldsymbol{\alpha}} \\ 0 &= 2[-\boldsymbol{B}^T\boldsymbol{y} + (\boldsymbol{B}^T\boldsymbol{B} + \lambda\boldsymbol{D_d}^T\boldsymbol{D_d})\hat{\boldsymbol{\alpha}}] \end{aligned} \tag{3.23}$$

Solving for $\hat{\boldsymbol{\alpha}}$ yields

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{B}^T\boldsymbol{B} + \lambda\boldsymbol{D_d}^T\boldsymbol{D_d})^{-1}\boldsymbol{B}^T\boldsymbol{y} \tag{3.24}$$

subsequently giving a *projection* or *hat* matrix of

$$\boldsymbol{H} = \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B} + \lambda\boldsymbol{D_d}^T\boldsymbol{D_d})^{-1}\boldsymbol{B}^T \tag{3.25}$$

3.2.2 2-Dimensional Case

The least squares term for the 2-dimensional B-spline function f can be expressed as

$$\sum_{i=1}^{n}[y_i - f(x_{1_i}, x_{2_i})]^2 \tag{3.26}$$

for ordered triplet (x_{1_i}, x_{2_i}, y_i). The 1-dimensional B-spline function involves a vector of coefficients. The 2-dimensional case has the coefficients (depicted in equation 3.7) as elements of a $k_1 \times k_2$ matrix, say $\boldsymbol{A} = [\alpha_{jl}]$ for $j = 1, ..., k_1$ and $l = 1, ..., k_2$, where k_1 and k_2 again represent the number knots in the B-spline bases for x_1 and x_2 respectively. Where the penalty for the 1-dimensional P-spline function takes first and higher order differences of a vector of coefficients, the 2-dimensional case takes the same differences

for every row and every column of $\boldsymbol{A}$. For example, the 2-dimensional P-spline penalty for difference order $d = 1$ and a common row and column penalty parameter can be expressed as

$$\lambda\left[\sum_{j=1}^{k_1}\sum_{l=2}^{k_2}(\alpha_{jl} - \alpha_{j(l-1)})^2 + \sum_{j=2}^{k_1}\sum_{l=1}^{k_2}(\alpha_{jl} - \alpha_{(j-1)l})^2\right] \tag{3.27}$$

where the first term takes differences of coefficients within the same row of $\boldsymbol{A}$ and the second term takes differences of coefficients within the same column of $\boldsymbol{A}$.

The next step is to calculate these row-wise and column-wise differences in a simple and efficient way. Let $\boldsymbol{\alpha}$ be a column vector composed of the $k_1 \times k_2$ elements of $\boldsymbol{A}$. The first k_1 elements of $\boldsymbol{\alpha}$ are the elements in the first row of $\boldsymbol{A}$, the second k_1 elements of $\boldsymbol{\alpha}$ are the elements in the second row of $\boldsymbol{A}$, etc... thus having the form

$$\boldsymbol{\alpha} = (\alpha_{11}, ..., \alpha_{1k_2}, \alpha_{21}, ..., \alpha_{2k_2}, ..., \alpha_{k_1 1}, ..., \alpha_{k_1 k_2})^T. \tag{3.28}$$

The differences of each set of k_2 elements of $\boldsymbol{\alpha}$ (the first k_2 elements, the second k_2 elements, etc...), corresponding to each of the k_1 rows of $\boldsymbol{A}$, can be calculated with the use of the Kronecker product

$$\boldsymbol{I_{k_1}} \otimes \boldsymbol{D_d}. \tag{3.29}$$

where $\boldsymbol{D_d}$ is a $(k_2 - d) \times k_2$ matrix. To illustrate the correct coefficients are being differenced, consider the example where the difference order $d = 1$ and $k_1 = k_2 = 4$. The coefficient matrix

$$\boldsymbol{A} = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} \end{bmatrix} \tag{3.30}$$

becomes the coefficient vector

$$\boldsymbol{\alpha} = (\alpha_{11}, \alpha_{12}, \alpha_{13}, \alpha_{14}, \alpha_{21}, \alpha_{22}, \alpha_{23}, \alpha_{24}, \alpha_{31}, \alpha_{32}, \alpha_{33}, \alpha_{34}, \alpha_{41}, \alpha_{42}, \alpha_{43}, \alpha_{44})^T \tag{3.31}$$

The Kronecker product of $\boldsymbol{I_4}$ with $\boldsymbol{D_1}$

$$\boldsymbol{I_4} \otimes \boldsymbol{D_1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \tag{3.32}$$

yields the 12×16 matrix

$$\begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}. \tag{3.33}$$

The multiplication $(\boldsymbol{I_4} \otimes \boldsymbol{D_1})\boldsymbol{\alpha}$ gives the vector of row-wise first differences of $\boldsymbol{A}$.

Furthermore, the differences of elements at the same position within each group of k_2 elements (the first element in the first k_2 elements, the first element in the second k_2 elements, etc...), corresponding to each of the k_2 columns of $\boldsymbol{A}$, can be calculated with the use of the Kronecker product

$$\boldsymbol{D_d} \otimes \boldsymbol{I_{k_2}} \tag{3.34}$$

where $\boldsymbol{D_d}$ is a $(k_1 - d) \times k_1$ matrix. Thus the Kronecker product of $\boldsymbol{D_1}$ with $\boldsymbol{I_4}$

$$\boldsymbol{D_1} \otimes \boldsymbol{I_4} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.35}$$

yields the 12×16 matrix

$$\begin{bmatrix}
-1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1
\end{bmatrix}. \tag{3.36}$$

The multiplication $(\boldsymbol{D_1} \otimes \boldsymbol{I_4})\boldsymbol{\alpha}$ gives the vector of column-wise first differences of $\boldsymbol{A}$.

The objection function to be minimized can now be generally expressed as

$$S = \|\boldsymbol{y} - \boldsymbol{B\alpha}\|^2 + \lambda \left[\left\| (\boldsymbol{I_{k_1}} \otimes \boldsymbol{D_d})\boldsymbol{\alpha} \right\|^2 + \left\| (\boldsymbol{D_d} \otimes \boldsymbol{I_{k_2}})\boldsymbol{\alpha} \right\|^2 \right] \tag{3.37}$$

for difference order d. In the interest of neat and concise notation, let

$$\boldsymbol{P_{row}} = (\boldsymbol{I_{k_1}} \otimes \boldsymbol{D_d})^T (\boldsymbol{I_{k_1}} \otimes \boldsymbol{D_d}), \tag{3.38}$$

$$\boldsymbol{P_{col}} = (\boldsymbol{D_d} \otimes \boldsymbol{I})^T (\boldsymbol{D_d} \otimes \boldsymbol{I}), \tag{3.39}$$

and

$$\boldsymbol{P_2} = \boldsymbol{P_{row}} + \boldsymbol{P_{col}}. \tag{3.40}$$

Following identity 3.20, the objective function can be expressed as

$$\begin{aligned}
S &= (\boldsymbol{y} - \boldsymbol{B\alpha})^T (\boldsymbol{y} - \boldsymbol{B\alpha}) + \lambda [\boldsymbol{\alpha}^T \boldsymbol{P_{row}} \boldsymbol{\alpha} + \boldsymbol{\alpha}^T \boldsymbol{P_{col}} \boldsymbol{\alpha}] \\
&= (\boldsymbol{y} - \boldsymbol{B\alpha})^T (\boldsymbol{y} - \boldsymbol{B\alpha}) + \lambda \boldsymbol{\alpha}^T \boldsymbol{P_2} \boldsymbol{\alpha} \\
&= \boldsymbol{y}^T \boldsymbol{y} - 2\boldsymbol{\alpha}^T \boldsymbol{B}^T \boldsymbol{y} + \boldsymbol{\alpha}^T \boldsymbol{B}^T \boldsymbol{B} \boldsymbol{\alpha} + \lambda \boldsymbol{\alpha}^T \boldsymbol{P_2} \boldsymbol{\alpha} \\
&= \boldsymbol{y}^T \boldsymbol{y} - 2\boldsymbol{\alpha}^T \boldsymbol{B}^T \boldsymbol{y} + \boldsymbol{\alpha}^T \Big(\boldsymbol{B}^T \boldsymbol{B} + \lambda \boldsymbol{P_2} \Big) \boldsymbol{\alpha}.
\end{aligned} \tag{3.41}$$

The estimate $\hat{\boldsymbol{\alpha}}$ is obtained by minimizing S as shown in equation 3.23. Solving for $\hat{\boldsymbol{\alpha}}$ yields

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{B}^T\boldsymbol{B} + \lambda \boldsymbol{P_{two}})^{-1}\boldsymbol{B}^T\boldsymbol{y} \tag{3.42}$$

subsequently giving the projection matrix

$$\boldsymbol{H} = \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B} + \lambda \boldsymbol{P_{two}})^{-1}\boldsymbol{B}^T \tag{3.43}$$

3.2.3 3-Dimensional Case

There is a natural extension from the 2-dimensional case to the 3-dimensional case. The least squares term can be expressed as

$$\sum_{i=1}^{n}[y_i - f(x_{1_i}, x_{2_i}, x_{3_i})]^2 \tag{3.44}$$

for ordered quadruples $(x_{1_i}, x_{2_i}, x_{3_i}, y_i)$. The coefficients depicted in equation 3.9 can be represented as elements of a $k_1 \times k_2 \times k_3$ 3-dimensional array $\boldsymbol{A} = [\alpha_{jlm}]$ where $j = 1, ..., k_1, l = 1, ..., k_2$, and $m = 1, ..., k_3$. Once again, k_1, k_2 and k_3 represent the number of B-spline bases knots for x_1, x_2, and x_3 respectively. The penalty now becomes

$$\begin{aligned}\lambda\Bigg[&\sum_{j=1}^{k_1}\sum_{l=2}^{k_2}\sum_{m=1}^{k_3}(\alpha_{jlm} - \alpha_{j(l-1)m})^2 + \\ &\sum_{j=2}^{k_1}\sum_{l=1}^{k_2}\sum_{m=1}^{k_3}(\alpha_{jlm} - \alpha_{(j-1)lm})^2 + \\ &\sum_{j=1}^{k_1}\sum_{l=1}^{k_2}\sum_{m=2}^{k_3}(\alpha_{jlm} - \alpha_{jl(m-1)})^2\Bigg]\end{aligned} \tag{3.45}$$

where the first term takes differences of coefficients within the same row of $\boldsymbol{A}$, the second term takes differences of coefficients within the same column of $\boldsymbol{A}$, and the third term takes differences of coefficients within the same slice Similar to the 2-dimensional case, the array of coefficients can be represented as the vector

$$\boldsymbol{\alpha} = (\alpha_{111}, ..., \alpha_{11k_3}, \alpha_{211}, ..., \alpha_{21k_3}, ..., \alpha_{k_1k_21}, ..., \alpha_{k_1k_2k_3})^T. \tag{3.46}$$

The row, column, and slice-wise d order differences of $\boldsymbol{A}$ can be calculated using the Kronecker products

$$\boldsymbol{K_{DII}} = \boldsymbol{D_d} \otimes \boldsymbol{I_{k_2}} \otimes \boldsymbol{I_{k_3}} \tag{3.47}$$

where $\boldsymbol{D_d}$ is a $(k_1 - d) \times k_1$ matrix,

$$\boldsymbol{K_{IDI}} = \boldsymbol{I_{k_1}} \otimes \boldsymbol{D_d} \otimes \boldsymbol{I_{k_3}} \tag{3.48}$$

where $\boldsymbol{D_d}$ is a $(k_2 - d) \times k_2$ matrix, and

$$\boldsymbol{K_{IID}} = \boldsymbol{I_{k_1}} \otimes \boldsymbol{I_{k_2}} \otimes \boldsymbol{D_d} \tag{3.49}$$

where $\boldsymbol{D_d}$ is a $(k_3 - d) \times k_3$ matrix, respectively. The objection function to be minimized can now be generally expressed as

$$S = \|\boldsymbol{y} - \boldsymbol{B\alpha}\|^2 + \lambda \left[\|\boldsymbol{K_{DII}\alpha}\|^2 + \|\boldsymbol{K_{IDI}\alpha}\|^2 + \|\boldsymbol{K_{IID}\alpha}\|^2 \right]. \tag{3.50}$$

Once again, in the interest of neat and concise notation, let

$$\boldsymbol{P_{row}} = \boldsymbol{K_{DII}}^T \boldsymbol{K_{DII}}, \tag{3.51}$$

$$\boldsymbol{P_{col}} = \boldsymbol{K_{IDI}}^T \boldsymbol{K_{IDI}}, \tag{3.52}$$

$$\boldsymbol{P_{sli}} = \boldsymbol{K_{IID}}^T \boldsymbol{K_{IID}}. \tag{3.53}$$

and

$$\boldsymbol{P_3} = \boldsymbol{P_{row}} + \boldsymbol{P_{col}} + \boldsymbol{P_{sli}}. \tag{3.54}$$

Following the 1 and 2-dimensional cases, the estimate $\hat{\boldsymbol{\alpha}}$ is given by

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{B}^T\boldsymbol{B} + \lambda \boldsymbol{P_3})^{-1} \boldsymbol{B}^T \boldsymbol{y} \tag{3.55}$$

subsequently giving the projection matrix

$$\boldsymbol{H} = \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B} + \lambda \boldsymbol{P_3})^{-1} \boldsymbol{B}^T \tag{3.56}$$

3.3 Additive Models

The River Run data set contains relationships between the response variable and the explanatory variables which are too complex to be represented by customary parametric models. Additive models provide an alternative to parametric models in these instances.

A simple additive model composed of only univariate P-spline functions can be represented by

$$y_i = \mu + \sum_{j=1}^{p} f_j(x_{j_i}) + \epsilon_i \tag{3.57}$$

where the y is the response variable, μ is the overall mean, x_j is the j^{th} explanatory variable, p is the number of explanatory variables, f_j it the smooth function corresponding to explanatory variable x_j, and the ϵ's are random noise, for $i = 1, ..., n$ observation. Each function f_j can be given a B-spline representation

$$f_j(x_{j_i}) = \sum_{r=1}^{k_j} B_{j_r}(x_{j_i})\alpha_{j_r} = \boldsymbol{B_j \alpha_j} \tag{3.58}$$

where

$$\boldsymbol{\alpha_j} = (\alpha_{j_1}, \alpha_{j_2}, ..., \alpha_{j_{k_j}})^T. \tag{3.59}$$

and where $\boldsymbol{B_j}$ is a block of $\boldsymbol{B^{(\mu:1)}}$, the full univariate B-spline matrix, which can be denoted as

$$\boldsymbol{B^{(\mu:1)}} = [\mathbf{1} : \boldsymbol{B_1} : \boldsymbol{B_2} : ... : \boldsymbol{B_p}] \tag{3.60}$$

Furthermore, equation 3.57 can be written in the vector form

$$\boldsymbol{y} = \boldsymbol{B^{(\mu:1)}\alpha^{(\mu:1)}} + \boldsymbol{\epsilon} \tag{3.61}$$

where

$$\boldsymbol{\alpha^{(\mu:1)}} = (\mu, \boldsymbol{\alpha_1}^T, \boldsymbol{\alpha_2}^T, ..., \boldsymbol{\alpha_p}^T)^T. \tag{3.62}$$

Here μ is the overall mean of the y_i's. The superscript $(\mu{:}1)$ aims to denote the mean and all univariate main effects are involved in the B-spline design matrix and corresponding coefficient vector. Although this notation may seem superfluous at this stage, it will make subsequent notation less complex. Furthermore, the subscripts on the α's become considerably more complicated as we move to bivariate interaction terms. To clarify, the notation here uses two levels of subscripts. The first level of subscripts, t and t', correspond to the index referencing the particular covariate of interest. The second level of subscripts, j and l, correspond to the particular B-spline within the basis for a specific covariate. An example is the coefficient $\alpha_{2,3_{4,6}}$ for the bivariate interaction term $f_{2,3}(x_2, x_3)$. This α corresponds to fourth B-spline within the B-spline basis of x_2 and

the sixth B-spline within the B-spline basis of x_3.

Additive models which include smoothed main effects of x_1 through x_p along with the bivariate smoothed interaction terms of every combination of explanatory covariates can be expressed as

$$y_i = \mu + \sum_{j=1}^{p} f_j(x_{j_i}) + \sum_{j=1}^{p-1} \sum_{l>j}^{p} f_{jl}(x_{j_i}, x_{l_i}) + \epsilon_i. \tag{3.63}$$

As in the univariate case with strictly main effects, bivariate interaction terms can be given by the B-spline representation

$$f_{jl}(x_{j_i}, x_{l_i}) = \sum_{r=1}^{k_j} \sum_{r'=1}^{k_l} B_{jl_{rr'}}(x_{j_i}, x_{l_i}) \alpha_{jl_{rr'}} = \boldsymbol{B_{jl}\alpha_{jl}} \tag{3.64}$$

where

$$\boldsymbol{\alpha_{jl}} = (\alpha_{jl_{1,1}}, \alpha_{jl_{1,2}}, ..., \alpha_{jl_{k_j k_l}})^T. \tag{3.65}$$

and where $\boldsymbol{B_{jl}}$, which involves covariates x_j and x_l, is a block of $\boldsymbol{B^{(2)}}$, the full bivariate B-spline matrix. Here

$$\begin{aligned} \boldsymbol{B^{(2)}} &= [\boldsymbol{B_{1,2}} : ... : \boldsymbol{B_{1,p}} : ... : \boldsymbol{B_{j,l}} : ... : \boldsymbol{B_{p-2,p-1}} : \boldsymbol{B_{p-2,p}} : \boldsymbol{B_{p-1,p}}] \\ &= [\boldsymbol{B_{j,l}}] \quad \forall\ j \text{ and } l \ni 1 \leq j < l \leq p \end{aligned} \tag{3.66}$$

in ascending numerical order and

$$\begin{aligned} \boldsymbol{\alpha^{(2)}} &= (\boldsymbol{\alpha_{1,2}}^T, ..., \boldsymbol{\alpha_{1,p}}^T, ..., \boldsymbol{\alpha_{j,l}}^T, ..., \boldsymbol{\alpha_{p-2,p-1}}^T, \boldsymbol{\alpha_{p-2,p}}^T, \boldsymbol{\alpha_{p-1,p}}^T)^T \\ &= ([\boldsymbol{\alpha_{j,l}}]^T)^T \quad \forall\ j \text{ and } l \ni 1 \leq j < l \leq p \end{aligned} \tag{3.67}$$

in ascending numerical order. To clarify "ascending numerical order", consider the example of $\boldsymbol{\alpha^{(2)}}$ where $p = 4$. Equation 3.67 would become

$$\boldsymbol{\alpha^{(2)}} = (\boldsymbol{\alpha_{1,2}}^T, \boldsymbol{\alpha_{1,3}}^T, \boldsymbol{\alpha_{1,4}}^T, \boldsymbol{\alpha_{2,3}}^T, \boldsymbol{\alpha_{2,4}}^T, \boldsymbol{\alpha_{3,4}}^T)^T. \tag{3.68}$$

The bivariate interaction terms, which will be incorporated into a model of the form 3.61, can be expressed as

$$\boldsymbol{B^{(2)}\alpha^{(2)}} \tag{3.69}$$

Finally, a three variable additive model with all possible interactions can be expressed as

$$y_i = \mu + \sum_{j=1}^{p} f_j(x_{j_i}) + \sum_{j=1}^{p-1}\sum_{l>j}^{p} f_{jl}(x_{j_i}, x_{l_i}) + \sum_{j=1}^{p-2}\sum_{l>j}^{p-1}\sum_{m>l}^{p} f_{jlm}(x_{j_i}, x_{l_i}, x_{m_i}) + \epsilon_i. \quad (3.70)$$

Following the bivariate interaction terms, the trivariate interaction terms can be represented by

$$\boldsymbol{B^{(3)}\alpha^{(3)}} \quad (3.71)$$

where

$$\begin{aligned} \boldsymbol{B^{(3)}} &= [\boldsymbol{B_{1,2,3}} : ... : \boldsymbol{B_{j,l,m}} : ... : \boldsymbol{B_{p-2,p-1,p}}] \\ &= [\boldsymbol{B_{j,l,m}}] \quad \forall\, j, l, \text{and } m \ni 1 \leq j < l < m \leq p \end{aligned} \quad (3.72)$$

and

$$\begin{aligned} \boldsymbol{\alpha^{(3)}} &= (\boldsymbol{\alpha_{1,2,3}}^T, ..., \boldsymbol{\alpha_{j,l,m}}^T, ..., \boldsymbol{\alpha_{p-2,p-1,p}}^T)^T \\ &= ([\boldsymbol{\alpha_{j,l,m}}]^T)^T \quad \forall\, j, l, \text{and } m \ni 1 \leq j < l < m \leq p \end{aligned} \quad (3.73)$$

arranged in ascending numerical order.

Essentially, $\boldsymbol{B^{(2)}}$ and $\boldsymbol{\alpha^{(2)}}$ both have $\binom{p}{2}$ elements $\boldsymbol{B_{j,l}}$ and $\boldsymbol{\alpha_{j,l}}$ respectively arranged in ascending numerical order and $\boldsymbol{B^{(3)}}$ and $\boldsymbol{\alpha^{(3)}}$ both have $\binom{p}{3}$ elements $\boldsymbol{B_{j,l,m}}$ and $\boldsymbol{\alpha_{j,l,m}}$ respectively arranged in ascending numerical order.

Combining univariate main effects with bivariate and trivariate interaction terms now allows us to express the additive model as

$$\boldsymbol{y = B\alpha + \epsilon} \quad (3.74)$$

where

$$\boldsymbol{B} = [\boldsymbol{B^{(\mu:1)}} : \boldsymbol{B^{(2)}} : \boldsymbol{B^{(3)}}] \quad (3.75)$$

and

$$\boldsymbol{\alpha} = ([\boldsymbol{\alpha^{(\mu:1)}}]^T, [\boldsymbol{\alpha^{(2)}}]^T, [\boldsymbol{\alpha^{(3)}}]^T)^T \quad (3.76)$$

The interaction terms can be interpreted as the adjustment to the response variable needed to account for the changing influence of one explanatory variable as the other

explanatory variables vary. The interaction terms can be extended to contain more explanatory covariates. Interaction terms of no more than three covariates will be implemented here.

3.3.1 Identifiability Constraints

Additive models containing multiple smooth functions have the inherent problem of identifiability; intercepts can not be estimated for all smooths. Therefore constraints can be put into place to address this issue. A simple but effective strategy is to force each main effect smooth to sum to zero for all p explanatory covariates and each interaction smooth to sum to zero and all combinations of p explanatory covariates respectively across each dimension.

3.3.1.1 Univariate Main Effects

In the case of an additive model containing only univariate main effects, the identifiablility constraint used here is

$$\sum_{i=1}^{n} \hat{f}_j(x_{j_i}) = 0 \quad \forall \quad j \in \{1, ..., p\}. \tag{3.77}$$

This is the method used in this research for the main effects. Higher order terms employ a different approach. Alternatively, Lee and Durbán (2011) and Ugarte et al. (2017) suggest, in the case of additive models composed of P-splines, the sum to zero constraint can be imposed on the basis coefficients;

$$\sum_{j=1}^{k_t} \alpha_{t_j} = 0 \quad \forall \quad t \in \{1, 2, 3, ..., p\}. \tag{3.78}$$

By using constraint equation 3.78 instead of 3.77, the computation becomes more efficient. For the 1-dimensional case the identifiability constraint requires

$$\mathbf{1_{k_t}}^T \boldsymbol{\alpha_t} = 0 \tag{3.79}$$

where

$$\boldsymbol{\alpha_t} = (\alpha_{t_1}, ...\alpha_{t_{k_t}})^T \tag{3.80}$$

and

$$\mathbf{1_{k_t}} = (1, ..., 1)^T \tag{3.81}$$

are both $k_t \times 1$ column vectors. The sum to zero constraint over each univariate main effect can be achieved by finding

$$min\left\{\left\|\mathbf{1_{k_t}^T \alpha_t}\right\|^2\right\} = min\left\{\boldsymbol{\alpha_t^T} \mathbf{1_{k_t}} \mathbf{1_{k_t}^T} \boldsymbol{\alpha_t}\right\} \quad \forall \quad t \in \{1, 2, 3, ..., p\}. \tag{3.82}$$

This allows the constraint to be implemented through an additional penalty. As $\boldsymbol{\alpha_t}$ is a vector of constants, expression 3.82 adjusts the α's so they are centered about the mean, μ, depicted in equation 3.62.

All p $\boldsymbol{\alpha_t}$'s can be summed to zero simultaneously by finding

$$min\left\{\left\|\boldsymbol{E_1 \alpha^{(1)}}\right\|^2\right\} = min\left\{\boldsymbol{\alpha^{(1)T} E_1}^T \boldsymbol{E_1 \alpha^{(1)}}\right\} \quad . \tag{3.83}$$

where $\boldsymbol{\alpha^{(1)}}$ is similar to equation 3.62 but with the lead μ removed, specifically

$$\boldsymbol{\alpha^{(1)}} = (\boldsymbol{\alpha_1}^T, \boldsymbol{\alpha_2}^T, ..., \boldsymbol{\alpha_p}^T)^T \tag{3.84}$$

and

$$\boldsymbol{E_1} = \begin{bmatrix} \mathbf{1_{k_1}}^T & k_2 \text{ zeroes} & \dots & k_{p-1} \text{ zeroes} & k_p \text{ zeroes} \\ k_1 \text{ zeroes} & \mathbf{1_{k_2}}^T & \dots & k_{p-1} \text{ zeroes} & k_p \text{ zeroes} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ k_1 \text{ zeroes} & k_2 \text{ zeroes} & \dots & \mathbf{1_{k_{p-1}}}^T & k_p \text{ zeroes} \\ k_1 \text{ zeroes} & k_2 \text{ zeroes} & \dots & k_{p-1} \text{ zeroes} & \mathbf{1_{k_p}}^T \end{bmatrix} \tag{3.85}$$

is a matrix composed of ones and zeroes of dimension

$$p \times \sum_{i=1}^{p} k_i. \tag{3.86}$$

3.3.1.2 Bivariate Interaction Terms

Additive models containing bivariate interaction terms must also have identifiability constraints imposed on those interaction terms. The sum to zero constraint on the

P-spline coefficients are

$$\sum_{j=1}^{k_t} \alpha_{tt'_{jl}} = 0 \quad \forall \quad t \text{ and } t' \in \{1, 2, 3, ..., p\} \quad \text{, where} \quad t \neq t', \text{ and}$$
$$\forall \quad l \in \{1, ..., k_l\} \tag{3.87}$$

and

$$\sum_{l=1}^{k_{t'}} \alpha_{tt'_{jl}} = 0 \quad \forall \quad t \text{ and } t' \in \{1, 2, 3, ..., p\} \quad \text{, where} \quad t \neq t', \text{ and}$$
$$\forall \quad j \in \{1, ..., k_j\}. \tag{3.88}$$

The condition of $t \neq t'$ in equations 3.87 and 3.88 assures the exclusion of interaction terms which do not have two distinct covariates, such as $f_{1,1}(x_1, x_1)$.

As each of the coefficients will be involved in two sum to zero constraints, a vector of strictly ones, such as $\mathbf{1_{k_t}}$, will not suffice. The 2-dimensional case requires the identifiability constraint

$$\mathbf{1_{k_t \cdot k_{t'}}} \boldsymbol{\alpha_{tt'}} = 0 \tag{3.89}$$

where

$$\boldsymbol{\alpha_{tt'}} = (\alpha_{tt'_{1,1}}, \alpha_{tt'_{1,2}}, ..., \alpha_{tt'_{1,k_{t'}}}, \alpha_{tt'_{2,1}}, ..., \alpha_{tt'_{2,k_{t'}}}, ..., \alpha_{tt'_{k_t,1}}, ..., \alpha_{tt'_{k_t,k_{t'}}})^T \tag{3.90}$$

is a $(k_t \cdot k_{t'}) \times 1$ column vector and

$$\mathbf{1_{k_t \cdot k_{t'}}} = \begin{bmatrix} \boldsymbol{I_{k_t}} \otimes \mathbf{1_{k_{t'}}}^T \\ \mathbf{1_{k_t}}^T \otimes \boldsymbol{I_{k_{t'}}} \end{bmatrix} \tag{3.91}$$

is a $(k_t + k_{t'}) \times (k_t \cdot k_{t'})$ matrix of ones and zeroes.

The sum to zero constraint over each bivariate interaction term can be achieved by finding

$$min\left\{ \left\| \mathbf{1_{k_t \cdot k_{t'}}} \boldsymbol{\alpha_{tt'}} \right\|^2 \right\} = min\left\{ \boldsymbol{\alpha_{tt'}}^T \mathbf{1_{k_t \cdot k_{t'}}}^T \mathbf{1_{k_t \cdot k_{t'}}} \boldsymbol{\alpha_{tt'}} \right\}$$
$$\forall \quad t \text{ and } t' \in \{1, 2, 3, ..., p\}, \text{where } t \neq t'. \tag{3.92}$$

All $\binom{p}{2}$ $\boldsymbol{\alpha_{tt'}}$'s can be summed to zero simultaneously by finding

$$min\left\{\left\|\boldsymbol{E_2\alpha^{(2)}}\right\|^2\right\} = min\left\{\boldsymbol{\alpha^{(2)}}^T\boldsymbol{E_2}^T\boldsymbol{E_2\alpha^{(2)}}\right\} \tag{3.93}$$

where $\boldsymbol{\alpha^{(2)}}$ is from equation 3.67 and

$$\boldsymbol{E_2} = \begin{bmatrix} \mathbf{1_{k_1 \cdot k_2}} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1_{k_1 \cdot k_3}} & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1_{k_{p-2} \cdot k_{p-1}}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{1_{k_{p-1} \cdot k_p}} \end{bmatrix} \tag{3.94}$$

is a matrix of ones and zeroes with dimension

$$\sum_{j=1}^{p-1}\sum_{l>j}^{p} k_j + k_l \times \sum_{j=1}^{p-1}\sum_{l>j}^{p} k_j \cdot k_l \tag{3.95}$$

3.3.1.3 Trivariate Interaction Terms

As is the case for univariate main effects and bivariate interaction terms, trivariate interaction terms of additive models must also have identifiability constraints imposed on their coefficients. The sum to zero constraints on the P-spline coefficients are

$$\sum_{j=1}^{k_t} \alpha_{tt't''_{jlm}} = 0 \quad \forall \quad t\,, t',\ \text{and}\ t'' \in \{1,2,3,...,p\} \quad \text{, where} \quad t \neq t' \neq t'' \neq t,\ \text{and}$$
$$\forall \quad l \in \{1,...,k_l\} \quad \text{and} \quad \forall \quad m \in \{1,...,k_m\} \tag{3.96}$$

and

$$\sum_{l=1}^{k_{t'}} \alpha_{tt't''_{jlm}} = 0 \quad \forall \quad t\,, t',\ \text{and}\ t'' \in \{1,2,3,...,p\} \quad \text{, where} \quad t \neq t'' \neq t' \neq t,\ \text{and}$$
$$\forall \quad j \in \{1,...,k_j\} \quad \text{and} \quad \forall \quad m \in \{1,...,k_m\} \tag{3.97}$$

and

$$\sum_{m=1}^{k_{t''}} \alpha_{tt't''_{jlm}} = 0 \quad \forall \quad t, t'\,, \text{and}\ t'' \in \{1,2,3,...,p\} \quad \text{, where} \quad t \neq t' \neq t'' \neq t,\ \text{and}$$
$$\forall \quad j \in \{1,...,k_j\} \quad \text{and} \quad \forall \quad l \in \{1,...,k_l\} \tag{3.98}$$

The 3-dimensional case requires the identifiability constraint

$$\mathbf{1}_{\boldsymbol{k_t}\cdot\boldsymbol{k_{t'}}\cdot\boldsymbol{k_{t''}}}\boldsymbol{\alpha_{tt't''}} = 0 \tag{3.99}$$

where

$$\boldsymbol{\alpha_{tt't''}} = (\alpha_{tt't''_{1,1,1}}, \alpha_{tt'_{1,1,2}}, ..., \alpha_{tt't''_{1,1,k_{t''}}}, \alpha_{tt't''_{1,2,1}}, ..., \alpha_{tt't''_{1,2,k_{t''}}}, ..., \alpha_{tt't''_{k_t,1,1}}, ..., \alpha_{tt't''_{k_t,k_{t'},k_{t''}}})^T \tag{3.100}$$

is a $(k_t \cdot k_{t'} \cdot k_{t''}) \times 1$ column vector and

$$\mathbf{1}_{\boldsymbol{k_t}\cdot\boldsymbol{k_{t'}}\cdot\boldsymbol{k_{t''}}} = \begin{bmatrix} \boldsymbol{I_{k_t}} \otimes \mathbf{1}_{\boldsymbol{k_{t'}}}{}^T \otimes \mathbf{1}_{\boldsymbol{k_{t''}}}{}^T \\ \mathbf{1}_{\boldsymbol{k_t}}{}^T \otimes \boldsymbol{I_{k_{t'}}} \otimes \mathbf{1}_{\boldsymbol{k_{t''}}}{}^T \\ \mathbf{1}_{\boldsymbol{k_t}}{}^T \otimes \mathbf{1}_{\boldsymbol{k_t}}{}^T \otimes \boldsymbol{I_{k_{t''}}} \end{bmatrix} \tag{3.101}$$

is a $(k_t + k_{t'} + k_{t''}) \times (k_t \cdot k_{t'} \cdot k_{t''})$ matrix of ones and zeroes.

The sum to zero constraint over each trivariate interaction term can be achieved by finding

$$min\left\{\left\|\mathbf{1}_{\boldsymbol{k_t}\cdot\boldsymbol{k_{t'}}\cdot\boldsymbol{k_{t''}}}\boldsymbol{\alpha_{tt't''}}\right\|^2\right\} = min\left\{\boldsymbol{\alpha_{tt't''}}{}^T\mathbf{1}_{\boldsymbol{k_t}\cdot\boldsymbol{k_{t'}}\cdot\boldsymbol{k_{t''}}}{}^T\mathbf{1}_{\boldsymbol{k_t}\cdot\boldsymbol{k_{t'}}\cdot\boldsymbol{k_{t''}}}\boldsymbol{\alpha_{tt't''}}\right\}$$
$$\forall \quad t, t'' \text{ and } t'' \in \{1, 2, 3, ..., p\}, \text{where } t \neq t' \neq t'' \neq t. \tag{3.102}$$

.

All $\binom{p}{3}$ $\boldsymbol{\alpha_{tt't''}}$'s can be summed to zero simultaneously by finding

$$min\left\{\left\|\boldsymbol{E_3}\boldsymbol{\alpha^{(3)}}\right\|^2\right\} = min\left\{\boldsymbol{\alpha^{(3)}}{}^T\boldsymbol{E_3}{}^T\boldsymbol{E_3}\boldsymbol{\alpha^{(3)}}\right\} \quad . \tag{3.103}$$

where $\boldsymbol{\alpha^{(3)}}$ is from equation 3.73 and

$$\boldsymbol{E_3} = \begin{bmatrix} \mathbf{1}_{\boldsymbol{k_1}\cdot\boldsymbol{k_2}\cdot\boldsymbol{k_3}}{}^T & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{\boldsymbol{k_1}\cdot\boldsymbol{k_2}\cdot\boldsymbol{k_4}}{}^T & \dots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{1}_{\boldsymbol{k_{p-2}}\cdot\boldsymbol{k_{p-1}}\cdot\boldsymbol{k_p}}{}^T \end{bmatrix} \tag{3.104}$$

is a matrix of ones and zeroes with dimension

$$\sum_{j=1}^{p-2}\sum_{l>j}^{p-1}\sum_{m>l}^{p} k_j + k_l + k_m \times \sum_{j=1}^{p-2}\sum_{l>j}^{p-1}\sum_{m>l}^{p} k_j \cdot k_l \cdot k_m. \tag{3.105}$$

3.3.1.4 Combining Main Effect and Interaction Identifiability Constraints

After characterizing the main effect and interaction identifiability constraints in this way, all these can now be handled collectively simply by adding additive penalty terms, one for each of these constraints This will ensure that the constraints are satisfied by locating the minimizing value of the parameters for each term. Alternatively, the sum to zero constraint over all main effects and interaction terms can be achieved by finding

$$min\left\{\left\|\boldsymbol{E\alpha^{(*)}}\right\|^2\right\} = min\left\{\boldsymbol{\alpha^{(*)T}E^TE\alpha^{(*)}}\right\} \quad . \tag{3.106}$$

where

$$\boldsymbol{\alpha^{(*)}} = (\boldsymbol{\alpha^{(1)T}}, \boldsymbol{\alpha^{(2)T}}, \boldsymbol{\alpha^{(3)T}})^T \tag{3.107}$$

is a vector of dimension

$$\left[p + \sum_{j=1}^{p-1}\sum_{l>j}^{p} k_j + k_l + \sum_{j=1}^{p-2}\sum_{l>j}^{p-1}\sum_{m>l}^{p} k_j + k_l + k_m\right] \times 1 \tag{3.108}$$

and

$$\boldsymbol{E} = \begin{bmatrix} \boldsymbol{E_1} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{E_2} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{E_3} \end{bmatrix} \tag{3.109}$$

is a matrix of ones and zeroes with dimension

$$\begin{aligned}&\left[p + \sum_{j=1}^{p-1}\sum_{l>j}^{p} k_j + k_l + \sum_{j=1}^{p-2}\sum_{l>j}^{p-1}\sum_{m>l}^{p} k_j + k_l + k_m\right] \times \\ &\left[\sum_{i=1}^{p} k_i + \sum_{j=1}^{p-1}\sum_{l>j}^{p} k_j \cdot k_l + \sum_{j=1}^{p-2}\sum_{l>j}^{p-1}\sum_{m>l}^{p} k_j \cdot k_l \cdot k_m\right].\end{aligned} \tag{3.110}$$

3.4 Additive Models With Random Effects for River Run Data

Natural groupings often arise in data collection. The sampling regime for the River Run data set has SEPA personnel collecting data with a substantial period between sample dates. There may exist unidentified factors driving the dissolved oxygen levels. These unidentified factors may individually or collectively cause a substantial variation in the dissolved oxygen from sampling date to sampling date. In each survey measurements at all the sampling stations are taken on the same day. They are therefore all subject to the same river conditions on that day. Adding a random effects component to the River Run model can be crucial to help explain the variation.

A mixed model is a model which combines fixed effect terms with random effect terms. A mixed model has the form

$$y_{ij} = f(x_{ij}) + \nu_j + \epsilon_{ij} \tag{3.111}$$

where y is the response variable, x is the explanatory variable, f is a fixed term (a smoothed function in this case), ν is a random effect for survey, i is the data point index, and j is the grouping index. This analysis will make use of an additive model with a random effects component or an additive mixed model. Wood (2017) provides excellent explanations for additive models, mixed models, and additive mixed models.

3.4.1 Schall's Algorithm

Schall (1991) outlines an algorithm for the estimation of fixed effects, random effects, and components of dispersion in generalized linear models, introduced by Fellner (1986) and Fellner (1987). This algorithm can be extended to additive models with random effects, since additive models can be thought of as complex linear models.

Consider the additive mixed model

$$\boldsymbol{y} = \boldsymbol{B\alpha} + \boldsymbol{U_1 b_1} + ... + \boldsymbol{U_c b_c} + \boldsymbol{\epsilon} \tag{3.112}$$

where $\boldsymbol{B}$ is a known B-spline design matrix, $\boldsymbol{\alpha}$ is vector of B-spline coefficients, the $\boldsymbol{U_i}$'s are known $n \times q_i$ matrices, and the $\boldsymbol{b_i}$'s are $q_i \times 1$ vectors of random effects for $i \in \{1, ..., c\}$. The random vectors $\boldsymbol{b_1}, ..., \boldsymbol{b_c}$ are assumed to be uncorrelated with zero expectation and uncorrelated with ϵ. Let

$$\boldsymbol{D} = \text{cov}(\boldsymbol{b}) = \text{diag}(\sigma_1^2 \boldsymbol{I_{q_1}}, ..., \sigma_c^2 \boldsymbol{I_{q_c}}) \tag{3.113}$$

where

$$\boldsymbol{b} = [\boldsymbol{b_1^T}, ..., \boldsymbol{b_c^T}]^T \tag{3.114}$$

Model 3.112 has $E(\boldsymbol{y}) = \boldsymbol{B\alpha}$, $\text{cov}(\boldsymbol{\epsilon}) = \boldsymbol{V} = \sigma^2\boldsymbol{I_n}$, and $\text{cov}(\boldsymbol{y}) = \boldsymbol{V} + \boldsymbol{UDU^T}$ where $\boldsymbol{U} = [\boldsymbol{U_1} : ... : \boldsymbol{U_c}]$.

The algorithm is outlined by Schall (1991) as follows:

- **Step 1**

Given estimates $\hat{\sigma}^2$ and $\hat{\sigma_1}^2$,...,$\hat{\sigma_c}^2$ for σ^2 and σ_1^2,...,σ_c^2 , compute estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{b_1}}, ..., \hat{\boldsymbol{b_c}}$ for $\boldsymbol{\alpha}$ and $\boldsymbol{b_1}, ..., \boldsymbol{b_c}$ as least squares to the set of overdetermined linear equations

$$\boldsymbol{C}\begin{bmatrix}\hat{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{b}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{V^{-1}B} & \boldsymbol{V^{-1}U}\\ \boldsymbol{0} & \boldsymbol{D^{-1}}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{\alpha}}\\ \hat{\boldsymbol{b}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{V^{-1}y}\\ \boldsymbol{0}\end{bmatrix}, \tag{3.115}$$

where $\boldsymbol{V}$ and $\boldsymbol{D}$ are evaluated at the current variance components. There are two versions of the second step; one yielding maximum likelihood estimates and the other yielding restricted maximum likelihood estimates.

- **Step 2** (Maximum Likelihood Estimates)

Let $\boldsymbol{T^*}$ be the inverse of the matrix formed by the last $q = q_1 + ... + q_c$ rows and columns of $\boldsymbol{C^TC}$, partitioned conformably with $\boldsymbol{D}$ as

$$\begin{bmatrix}\boldsymbol{T_{11}^*} & \cdots & \boldsymbol{T_{1c}^*}\\ \vdots & & \vdots\\ \boldsymbol{T_{c1}^*} & \cdots & \boldsymbol{T_{cc}^*}\end{bmatrix} \tag{3.116}$$

Given estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{b_1}}, ..., \hat{\boldsymbol{b_c}}$ for $\boldsymbol{\alpha}$ and $\boldsymbol{b_1}, ..., \boldsymbol{b_c}$, compute estimates as $\hat{\sigma}^2$ and $\hat{\sigma_1}^2$,...,$\hat{\sigma_c}^2$ for σ^2 and σ_1^2,...,σ_c^2 as

$$\hat{\sigma}^2 = \frac{(\boldsymbol{y} - \boldsymbol{B\hat{\alpha}} - \boldsymbol{U\hat{b}})^T(\boldsymbol{y} - \boldsymbol{B\hat{\alpha}} - \boldsymbol{U\hat{b}})}{n - \sum_{i=1}^{c}(q_i - v_i^*)} \tag{3.117}$$

and

$$\hat{\sigma}_i^2 = \frac{\hat{b}_i^T\hat{b}_i}{q_i - v_i^*} \tag{3.118}$$

where $v_i^* = \text{trace}(\boldsymbol{T_{ii}^*})/\sigma_i^2$ is evaluated at the current estimate of σ_i^2.

- **Step 2** (Restricted Maximum Likelihood Estimates)

Let $\boldsymbol{T}$ be the matrix formed by the last $q = q_1 + ... + q_c$ rows and columns of $\boldsymbol{C^T C}$, partitioned conformably with $\boldsymbol{D}$ as

$$\begin{bmatrix} \boldsymbol{T_{11}} & \dots & \boldsymbol{T_{1c}} \\ \vdots & & \vdots \\ \boldsymbol{T_{c1}} & \dots & \boldsymbol{T_{cc}} \end{bmatrix} \tag{3.119}$$

Given estimates $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{b_1}}, ..., \hat{\boldsymbol{b_c}}$ for $\boldsymbol{\alpha}$ and $\boldsymbol{b_1}, ..., \boldsymbol{b_c}$, compute estimates $\hat{\sigma}^2$ and $\hat{\sigma_1}^2$,...,$\hat{\sigma_c}^2$ for σ^2 and σ_1^2,...,σ_c^2 as

$$\hat{\sigma}^2 = \frac{(\boldsymbol{y} - \boldsymbol{B\hat{\alpha}} - \boldsymbol{U\hat{b}})^T(\boldsymbol{y} - \boldsymbol{B\hat{\alpha}} - \boldsymbol{U\hat{b}})}{n - \sum_{i=1}^{c}(q_i - v_i)} \tag{3.120}$$

and

$$\hat{\sigma}_i^2 = \frac{\hat{b}_i^T \hat{b}_i}{q_i - v_i} \tag{3.121}$$

where $v_i = \text{trace}(\boldsymbol{T_{ii}})/\sigma_i^2$ is evaluated at the current estimate of σ_i^2.

Schall (1991) continues and adapts the algorithm for estimation in generalized linear models with random effects, but this is not relevant to this thesis.

3.5 Controlling Smoothing

When fitting a smooth model, one needs to consider how much smoothing is appropriate. Selecting a large smoothing parameter will lead to over-smoothing and subsequently elevates the bias. Choosing a smoothing parameter too small will cause over-fitting causing the variance to increase. A variety of methods of selecting the degree of smoothing have been proposed over the years. The common criteria are generalized cross-validation (GCV), Akaike's information criterion (AIC), corrected Akaike's information criterion (AICc), and Bayesian information criterion. In each case the criteria balances model fit against model complexity. The penalty parameter is chosen to minimize the chosen criteria. Some useful references in the context of P-splines are Eilers and Marx (1992), Lee (2003), Wahba et al. (1985), Ruppert (2002), and Imoto and Konishi (2003).

Wood et al. (2015), when addressing generalized additive models for large data sets, gives the generalized cross-validation score as

$$GCV = \frac{n \cdot \sum_{i=1}^{n}[y_i - \hat{y}(x_i)]^2}{[n - tr(\boldsymbol{H})]^2}. \tag{3.122}$$

where $tr(\boldsymbol{H})$ is the trace of the hat or projection matrix which estimates the degrees of freedom of the model.

First introduced by Akaike (1973), AIC is given by

$$AIC = 2 \cdot tr(\boldsymbol{H}) + n \cdot log\left(\frac{\sum_{i=1}^{n}[y_i - \hat{y}(x_i)]^2}{n}\right). \tag{3.123}$$

deLeeuw (1992) mentioned it was more original and daring than cross-validation.

First introduce by Sugiura (1978), AICc is expressed as

$$AICc = n \cdot log\left(\frac{\sum_{i=1}^{n}[y_i - \hat{y}(x_i)]^2}{n}\right) + 1 + 2 \cdot \left(\frac{tr(\boldsymbol{H}) + 1}{n - tr(\boldsymbol{H}) - 2}\right). \tag{3.124}$$

AICc is asymptotically efficient in regression. For linear regression, AICc is exactly unbiased as long as candidate family of models includes the true model. For nonlinear regression, the unbiasedness is only approximate. The reduction in bias is achieved without any increase in variance, since AICc may be written as the sum of AIC and a nonstochastic term (Hurvich and Tsai (1989)).

Introduced by Schwarz et al. (1978), BIC is given by

$$BIC = log(n) \cdot tr(\boldsymbol{H}) + n \cdot log\left(\frac{\sum_{i=1}^{n}[y_i - \hat{y}(x_i)]^2}{n}\right). \tag{3.125}$$

Xue and Yang (2006) find model selection based on BIC outperforms AIC when applied to models with i.i.d. errors and models with a time series set-up. AIC tended to overfit in both cases. Chakrabarti and Ghosh (2011) compared AIC and BIC in several examples and found, while AIC generally chose a model with better predicting power, BIC did a better job of selecting the correct model.

3.6 Summary

Chapter 3 details the methodology for constructing P-spline additive models with interaction which will be used in this thesis. B-splines are introduced with their general properties and defined recursively. The structure of univariate, bivariate, and trivariate

B-splines are described. Examples of B-splines and B-spline bases of order 1 through 4 are depicted graphically for univariate and bivariate functions. P-splines are then introduced with an explanation of how a penalty involving differences on the B-spline coefficients is implemented for univariate functions. Kronecker products are used to implemented difference penalties on bivariate and trivariate functions. The B-spline coefficients are then estimated with these penalties in place.

The structure of P-spline additive models is detailed by combining univariate, bivariate, and trivariate B-spline matrices in a block matrix with the coefficients also combined in a block vector. Identifiability constraints are detailed by using Kronecker products. Schall's algorithm for fitting the random effects is explained. Chapter 3 finishes by describing the various scores which will be used for smoothing parameter selection.

With the methodology of the various aspects of fitting an additive mixed model with interactions detailed, the next stage of this thesis will address the analysis of variance with additive model terms. Although this thesis will not address the analysis of variance of additive mixed model terms, there is value in a simulation study checking the performance of various methods in detecting the presence of an interaction for additive models with no random effect component.

Chapter 4

Analysis of Variance with Additive Model Terms

When fitting additive models, it can be valuable to assess the evidence that particular model terms are needed. The probability of correctly detecting effects which are present (power) and the probability of correctly not detecting effects which are not present (size) can be assessed by simulation. Nested additive models, which include not only main effects but variable interaction terms, provide the motivation for this simulation study. The methods of model selection considered in this simulation study for the `sm` package are a simple F-test, an F-test involving quadratic forms, and a parametric bootstrap involving a simple F-test. How well these tests are calibrated and computational expense for each will be assessed. A comparison will be made with the `gam` function of the `mgcv` package created by Wood (2005) fitting comparable models and comparable model selection methods. Although mgcv is a powerful package for fitting additive mixed models, a locally constructed `sm` package was used. This allowed greater control over the details of the implementation of P-spline additive models, and the application of new methods of analysis. The inclusion of `mgcv` in this simulation study provides a comparison of performance against the methods implemented in `sm` in detecting the presence of interactions. Simple models involving no more than three explanatory covariates with bivariate interactions will be used.

4.1 Methods

All three methods considered in this simulation study involve the generation of the response variable from a model we will assume to be the *truth*. The *true* model will

have the form

$$y_i = \sum_{j=1}^{u} f_j(*) + \sum_{k=u+1}^{u+b} f_k(* : *) + \epsilon_i \tag{4.1}$$

where the y's represent the response variable,the f's represent smooth functions, u represents the number of univariate terms, b the number of bivariate terms, and ϵ is the i.i.d. random error for i= 1 to n, where n represents the number of observations. The *'s represent the explanatory covariates of the i^{th} observation corresponding to the particular smooth function f .

A specific term in an additive model can be expressed as

$$\widetilde{\boldsymbol{A}}\tilde{\boldsymbol{\alpha}} \tag{4.2}$$

where $\widetilde{\boldsymbol{A}}$ contains the columns of the full design matrix $\boldsymbol{B}$ which correspond to the term of interest and $\tilde{\boldsymbol{\alpha}}$ contains the coefficients of the full coefficient vector $\boldsymbol{\alpha}$ which correspond to the term of interest. A test can then be based on whether $\tilde{\boldsymbol{\alpha}} = 0$ by examining

$$\frac{\widehat{\tilde{\boldsymbol{\alpha}}}^T Var(\widehat{\tilde{\boldsymbol{\alpha}}})\widehat{\tilde{\boldsymbol{\alpha}}}}{\hat{\sigma}^2}. \tag{4.3}$$

4.1.1 Simple F-test

Consider the additive model $M1$ of the form

$$y_i = f_1(x_{1_i}) + f_2(x_{2_i}) + f_3(x_{1_i} : x_{2_i}) + \epsilon_i \tag{4.4}$$

and the additive model $M0$ of the form

$$y_i = f_1(x_{1_i}) + f_2(x_{2_i}) + \epsilon_i \tag{4.5}$$

for $i = 1$ to n, the x's are the explanatory variables,. The residual sum of squares is defined by

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}((x_i))^2. \tag{4.6}$$

Let df denote the degrees of freedom. The F test statistic is then defined by

$$F = \frac{(RSS_{M0} - RSS_{M1})/(df_{M1} - df_{M0})}{RSS_{M1}/(n - df_{M1})} \tag{4.7}$$

where the subscripts represent the model. The test statistic is to be compared to the F-distribution value corresponding degrees of freedom $(df_{M1} - df_{M0})$ and $(n - df_{M1})$. A version of the above equation in terms of degrees of freedom of each model error can be found page 80 of Bowman and Azzalini (1997). We are then left with the task of finding the probability of getting an F value greater than F_{obs}, the value calculated from the data. In other words finding p where

$$p = \mathbb{P}\{F > F_{obs}\}. \tag{4.8}$$

This method involves fitting both models $M1$ and $M0$. The same results can be attained by only fitting model $M1$ and using expression 4.3. Let $\tilde{\boldsymbol{V}}$ be the covariance matrix of $\hat{\tilde{\boldsymbol{\alpha}}}$. An equivalent F test statistic to the one in equation 4.7 is

$$F = \frac{\hat{\tilde{\boldsymbol{\alpha}}}^T \tilde{\boldsymbol{V}}^{-1} \hat{\tilde{\boldsymbol{\alpha}}}}{df_{\hat{\tilde{\boldsymbol{\alpha}}}}} \tag{4.9}$$

This test statistic is to be compared to the F-distribution value corresponding degrees of freedom $df_{\hat{\tilde{\boldsymbol{\alpha}}}}$ and $(n - df_{M1})$.

4.1.2 Quadratic Forms

The smoothing matrix $\boldsymbol{S}$ is defined by the equation

$$\hat{\boldsymbol{y}} = \boldsymbol{S}\boldsymbol{y} \tag{4.10}$$

where $\hat{\boldsymbol{y}}$ is a vector of predicted values and $\boldsymbol{y}$ is a vector of observed values (page 72) Bowman and Azzalini (1997). Thus S is analogous to the projection (or hat) matrix of the form $\boldsymbol{H}$ in equation 3.43. Let $\boldsymbol{S_{M1}}$ be the smoothing matrix of model $M1$. In testing for no effect in parametric regression, the *pseudo-likelihood ratio test* statistic

$$F = \frac{RSS_0 - RSS_1}{RSS_1} \tag{4.11}$$

is proportional to the F statistic in the previous section. This F statistic can be expressed in quadratic forms as

$$F = \frac{\boldsymbol{y}^T \boldsymbol{Q} \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{U} \boldsymbol{y}} \tag{4.12}$$

where

$$\boldsymbol{U} = (\boldsymbol{I} - \boldsymbol{S_{M1}})^T (\boldsymbol{I} - \boldsymbol{S_{M1}}) \tag{4.13}$$

and

$$\boldsymbol{Q} = \boldsymbol{B}\boldsymbol{G}^T\widetilde{\boldsymbol{A}}^T(\widetilde{\boldsymbol{A}}\boldsymbol{G}\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{G}^T\widetilde{\boldsymbol{A}}^T)^{-1}\widetilde{\boldsymbol{A}}\boldsymbol{G}\boldsymbol{B}^T \tag{4.14}$$

and

$$\boldsymbol{G} = (\boldsymbol{B}^T\boldsymbol{B} + \lambda\boldsymbol{P})^{-1} \tag{4.15}$$

where $\boldsymbol{P}$ is the penalty matrix. We are interested in finding

$$p = \mathbb{P}\{F > F_{obs}\} \tag{4.16}$$

which can be re-expressed as

$$\begin{aligned} p &= \mathbb{P}\left\{\frac{\boldsymbol{y}^T\boldsymbol{Q}\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{U}\boldsymbol{y}} > F_{obs}\right\} \\ &= \mathbb{P}\{\boldsymbol{y}^T\boldsymbol{C}\boldsymbol{y} > 0\} \end{aligned} \tag{4.17}$$

where

$$\boldsymbol{C} = (\boldsymbol{Q} - F_{obs}\boldsymbol{U}). \tag{4.18}$$

Since p values of three significant figures are sufficient for our purposes and the full calculations can become awkward, a scaled and shifted χ^2 distribution

$$a\chi^2_b + c \tag{4.19}$$

can be used by matching the first three or four moments of $\boldsymbol{\epsilon}^T\boldsymbol{C}\boldsymbol{\epsilon}$, which is equivalent to $\boldsymbol{y}^T\boldsymbol{C}\boldsymbol{y}$ because the differences in equation 4.18 cause the deterministic portions to drop out. These cumulants of $\boldsymbol{\epsilon}^T\boldsymbol{C}\boldsymbol{\epsilon}$ have the form

$$\kappa_j = 2^{j-1}(j-1)!tr[(\boldsymbol{V}\boldsymbol{C})^j], \tag{4.20}$$

where tr(*) represents the trace operator and $\boldsymbol{V}$ represents the covariance matrix of $\boldsymbol{y}$. Subsequently the values of a, b, and c become

$$a = |\kappa_3|/(4\kappa_2), \qquad b = 8(\kappa_2^3)/\kappa_3^2, \qquad c = \kappa_1 - ab. \tag{4.21}$$

which can be found on page 88 of Bowman and Azzalini (1997).

4.1.3 Parametric Bootstrap

Where the non-parametric bootstrap re-samples from the original data and makes no assumption of the underlying relation between explanatory and response variables in model selection, the parametric bootstrap for model selection assumes a specific relation. Randomly generated values of the explanatory variables are used to generate outputs based on a particular model. These outputs have random noise added to them to simulate a data set. Our goal is to check the performance of the bootstrap in correctly detecting influences which are present as well as correctly not detecting influences which are absent.

To better understand how the parametric bootstrap works, consider the following example. Consider once again models $M0$

$$y_i = f_1(x_{1_i}) + f_2(x_{2_i}) + \epsilon_i \tag{4.22}$$

and $M1$

$$y_i = f_1(x_{1_i}) + f_2(x_{2_i}) + f_3(x_{1_i} : x_{2_i}) + \epsilon_i \tag{4.23}$$

The simple F test statistic is calculated for the presence of the interaction term $f_3(x_1 : x_2)$. Denote this statistic by F_{M1}. The following steps are then carried out N_{boot} times.

1. Outputs, y_{boot}, generated by fitted model $M0$ with added random normal noise centered at 0 and variance corresponding to the fitted $M0$.
2. An $M1$ type model is then fit to y_{boot}, x_1, and x_2.
3. The simple F statistic is calculated for the presence of f_3 which we will denote by F_{boot}.

Once these three steps have been carried out N_{boot} times, a p value is calculated using the equation

$$p = \frac{\text{number of times } F_{boot} \text{ is greater than } F_{M_1}}{N_{boot}} \tag{4.24}$$

4.1.4 Comparing to mgcv

Since there exist R packages involving additive models, specifically mgcv, there is interest in seeing how the model selection methods therein perform compared to `sm`. The three model selection methods used here are listed below:

- F-ratio statistic called by the anova.gam function. It is used to test the null hypothesis that the simpler model is correct against the complex model (Wood (2017)).

- Chisq statistic called by the anova.gam function. It is used for assessing the significance of a particular smoothed term. If $\boldsymbol{p_i}$ is the parameter vector of the i^{th} smooth term, and this term has estimated covariance matrix $\boldsymbol{V_i}$, then the statistic is $\boldsymbol{p_i}^T \boldsymbol{V_i}^{k-} \boldsymbol{p_i}$, where $\boldsymbol{V_i}^{k-}$ is the rank $k-1$ pseudo-inverse of $\boldsymbol{V_i}$, and k is the basis dimension Wood (2005).

- SPV statistic called by the summary.gam function. It is a p-value for the null hypothesis that each smooth term is zero. The R documentation warns that these values are only approximate (Wood (2005)).

Although the degrees of freedom are adjusted to match those used for `sm`, there are inherent differences in the ways the models are fitted and the model selection methods for `sm` and `mgcv`. For this reason `mgcv` is only included in this simulation study as a check on `sm`. As a consequence there is also an opportunity to see the differences in model selection methods within `mgcv`.

4.2 Design

The primary purpose of this simulation study is to check the performance of the three different `sm` methods for detecting the presence of interaction terms in an additive model environment. A secondary aim is to compare the `sm` function methods with the `gam` function methods. To do this we must insure that each term in `sm` model has the same degrees of freedom as the corresponding term in the `gam` model. This was accomplished by adjusting the smoothing parameter of the `gam` model terms to attain degrees of freedom values equal to those of the `sm` model.

Since additive models are made up of smooth functions it seems appropriate to utilize functions with several combinations of linear, trigonometric, power, and exponential functions as main effects and interaction terms. The functions used in this study are

- **F1**: $y = 2cos(2\pi x_1) + 2x_2 + ae^{x_1 x_2}$
- **F2**: $y = 2sin(2\pi x_1) + 2e^{x_2} + a(x_1 x_2)^2$
- **F3**: $y = 2sin(2\pi x_1) + 2e^{x_2} + 2(x_3)^2 + a(x_1 x_2)^2$
- **F4**: $y = 2sin(2\pi x_1) + 2e^{x_2} + 10(x_3)^3 + a\sqrt{x_1 x_2}$

- **F5**: $y = 2sin(2\pi x_1) + 2e^{x_2} + 2cos(4\pi x_3) + a\sqrt{x_2 x_3}$
- **F6**: $y = 2sin(2\pi x_1) + 2e^{-x_2} + 10(x_3)^3 + a\sqrt{x_2 x_3}$
- **F7**: $y = 2sin(2\pi x_1) + 2e^{-x_2} + 10(x_3)^3 + a\sqrt{x_1 x_2}$
- **F8**: $y = 2sin(2\pi x_1) + 2e^{-x_2} + 10(x_3)^3 + a(sin(2\pi x_1)sin(2\pi x_2))$

where a scales the interaction term and takes on the values {0.2,0.5,0.7,1,1.5,2,2.5,3,3.5,5} and the x's are random uniform values from the interval [0,1].

When a is set to 0, representing the absence of the interaction term being tested, the p values for each of the three test methods should be uniformly distributed between 0 and 1. That is to say the test F statistic essentially has the potential of landing anywhere along the F distribution (or the shifted and scaled χ^2 distribution in the case of the quadratic forms test) with equal probability. As a becomes more and more positive the p values should tend towards values below to 0.05. Specifically we are interested in the proportion of p values less than or equal to 0.05. In the interest of minimizing computational expense in the future, the simulations were run under three scenarios:

1. 500 simulations with 500 bootstrap iterations within each simulation ($N_{sim} = 500$ and $N_{boot} = 500$)
2. 500 simulations with 200 bootstrap iterations within each simulation ($N_{sim} = 500$ and $N_{boot} = 200$)
3. 200 simulations with 200 bootstrap iterations within each simulation ($N_{sim} = 200$ and $N_{boot} = 200$).

If all three scenarios give similar results and the parametric bootstrap outperforms the other two `sm` methods, scenario 3 would save considerable computational time.

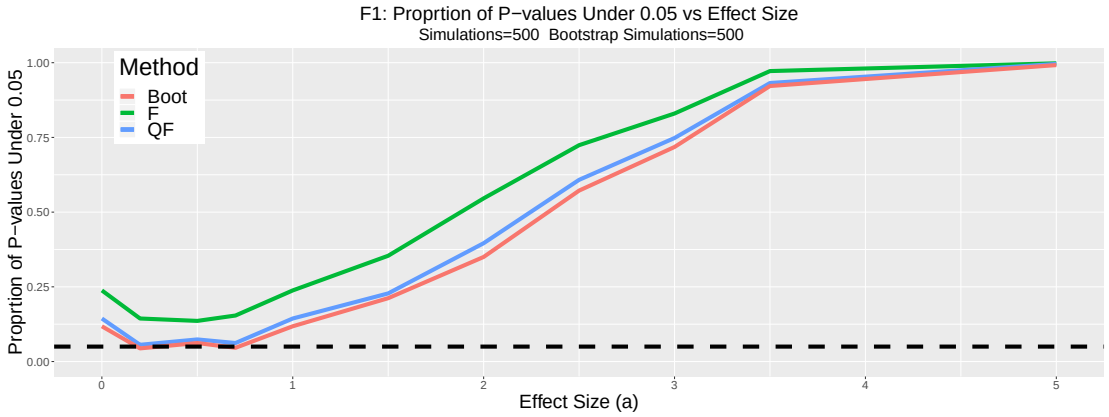
4.3 Results

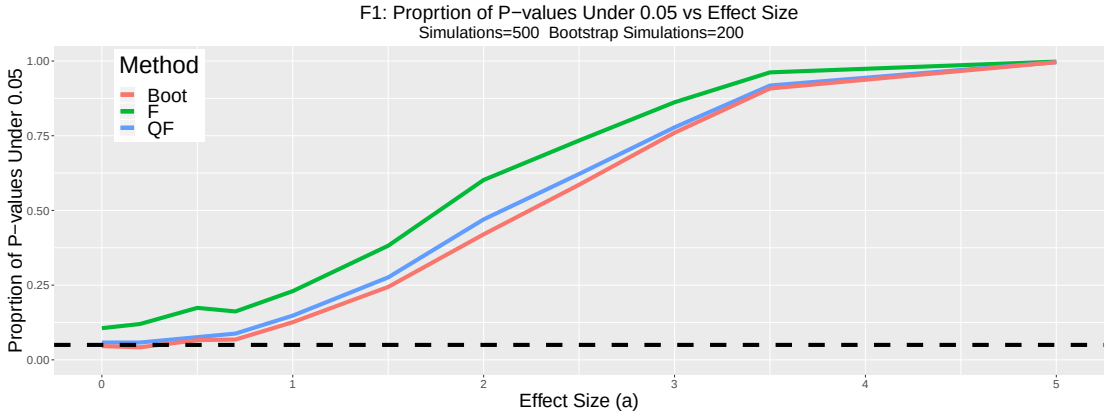
An unlikely event has been deemed to occur less than 5% of the time and so it is commonly the assigned level of significance. Figures 4.1-4.8 depict the proportion of p-values under 5% against the effect size (a). The panels from top to bottom in each figure represent scenarios 1, 2, 3, and `mgcv` respectively. Each method is represented by different colour lines for both sm and `mgcv`. The horizontal black dashed line represents 5%. This line helps to assess the probability the method will correctly not detect an effect which is not present (size). The probability the method will correctly detect an

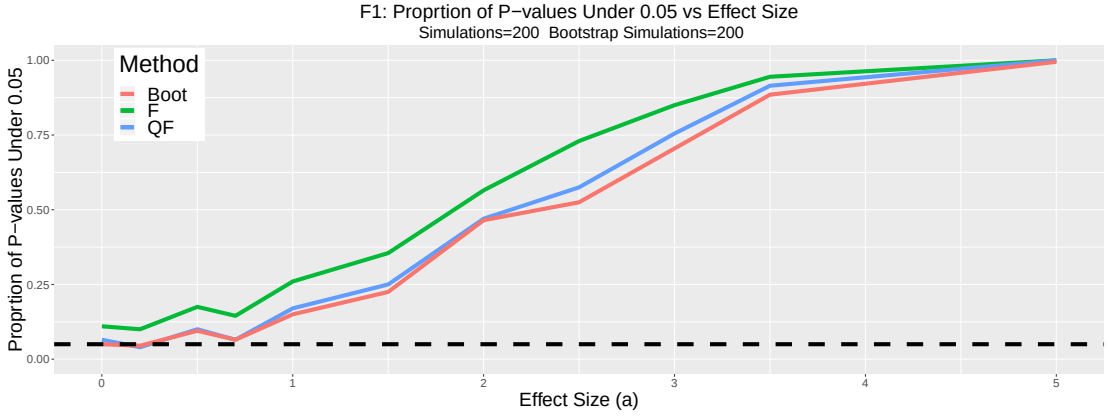
effect which is there (power) can be assessed by the increase in the coloured lines as the effect becomes stronger. Ideally, a well calibrated method will have a line that starts at or below the 5% dashed line for an effect size of 0 and increases above 5% consistently as the effect size becomes greater.

Upon reviewing Figures 4.1-4.8, one can see that for any one function, the 3 scenarios for `sm` do not differ much. Therefore, if the parametric bootstrap is to be used, scenario 3 with 200 bootstrap simulations would reduce computational expense. The F test for `sm` (green line) consistently starts off above the black dashed line showing it is falsely detecting an effect which is not present more than it should. However, the bootstrap (red line) and quadratic forms (blue line) tests for `sm` almost always start off on or below the black dashed line. This suggests the latter methods have adequate size whereas the former does not. As for power, the quadratic forms method is quicker to detect an effect in the presence of one. This is evident because the quadratic forms line is consistently on or above the bootstrap line. This is fortunate since the quadratic forms test is not nearly as computationally demanding as the bootstrap. Overall, this simulation study provides compelling evidence that the quadratic forms method of model selection should be the choice of preference for additive models but not necessarily useful for additive mixed models.

As mentioned in the beginning of Chapter 4, the comparison of `sm` to `mgcv` allowed the comparison of new methods in the detection of the presence of interactions. It seems in all but one function, **F8**, the `sm` methods were able to detect the presence of an interaction more efficiently and thus had better power. In fact `mgcv` was not able to detect the increasing interactions in functions **F5** and **F6**. Of the methods of ANOVA within `mgcv`, the F-test seems to have the superior size and power combination. The F-test consistently starts below the 5% dashed line when no interaction is present, unlike Chisq, and detects the interaction better than SPV. Although the computational cost is greater for `sm` when compared to `mgcv`, the methods implemented in the former offers the smoothing parameter control and the superior power in detecting the presence of interactions.







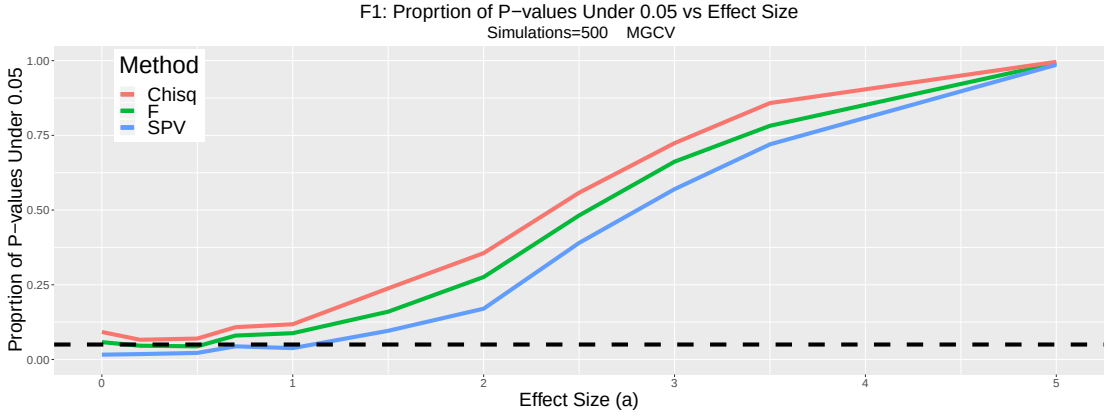
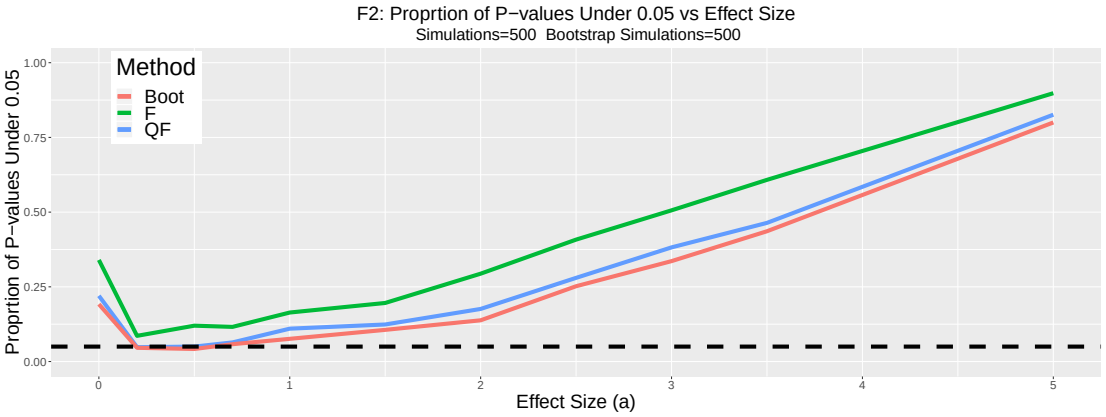
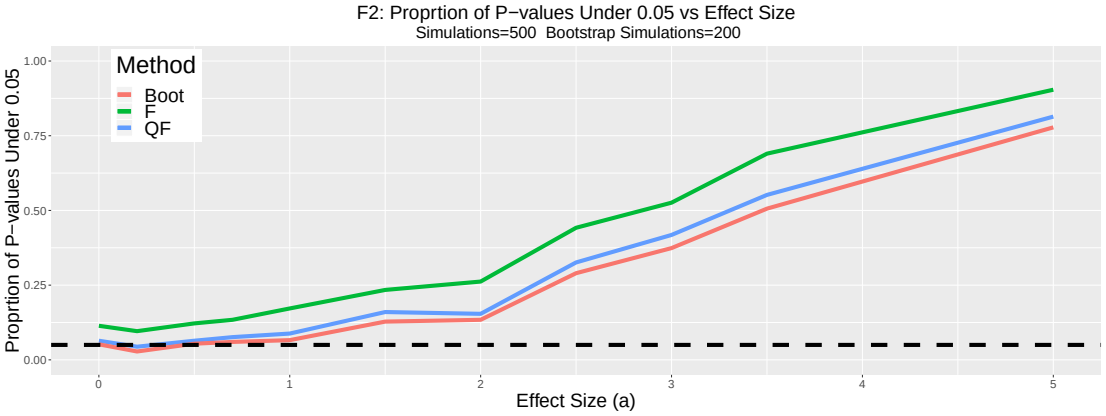
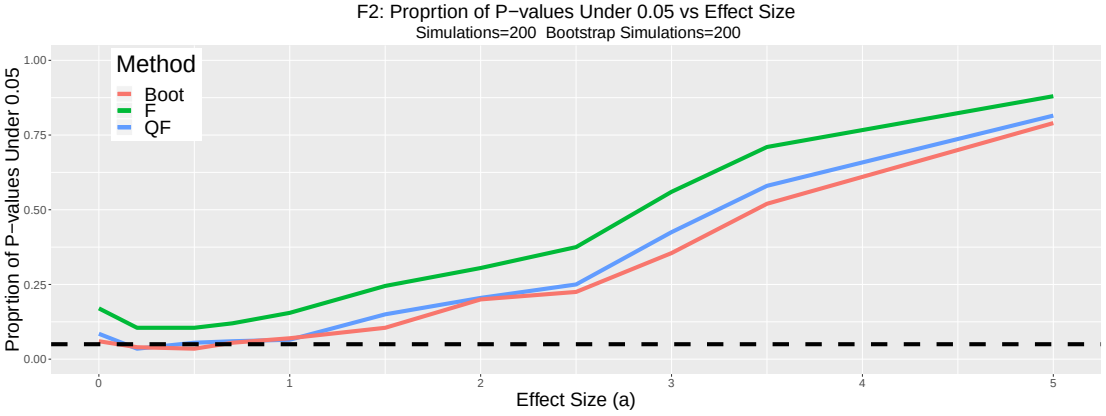


FIGURE 4.1: Plots showing the proportion of p-values under 5 % as a function of effect size for F1 for the 3 scenarios and `mgcv`.







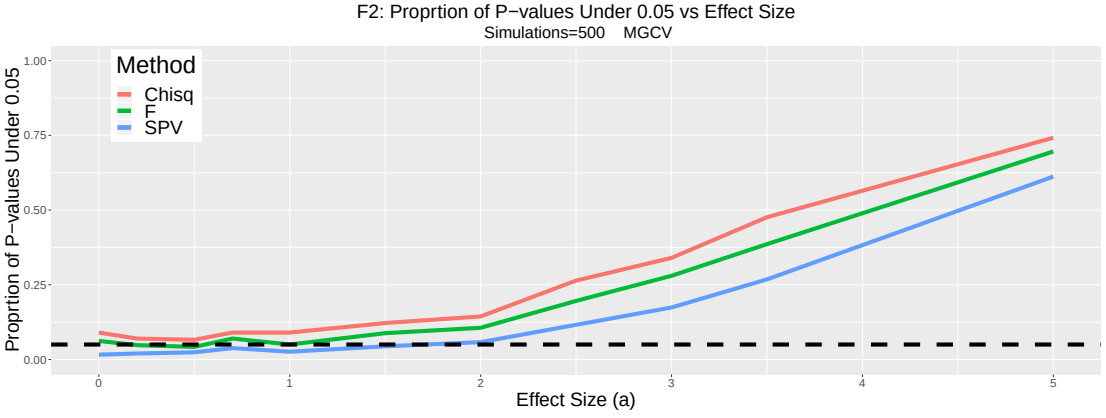


FIGURE 4.2: Plots showing the proportion of p-values under 5 % as a function of effect size for F2 for the 3 scenarios and `mgcv`.

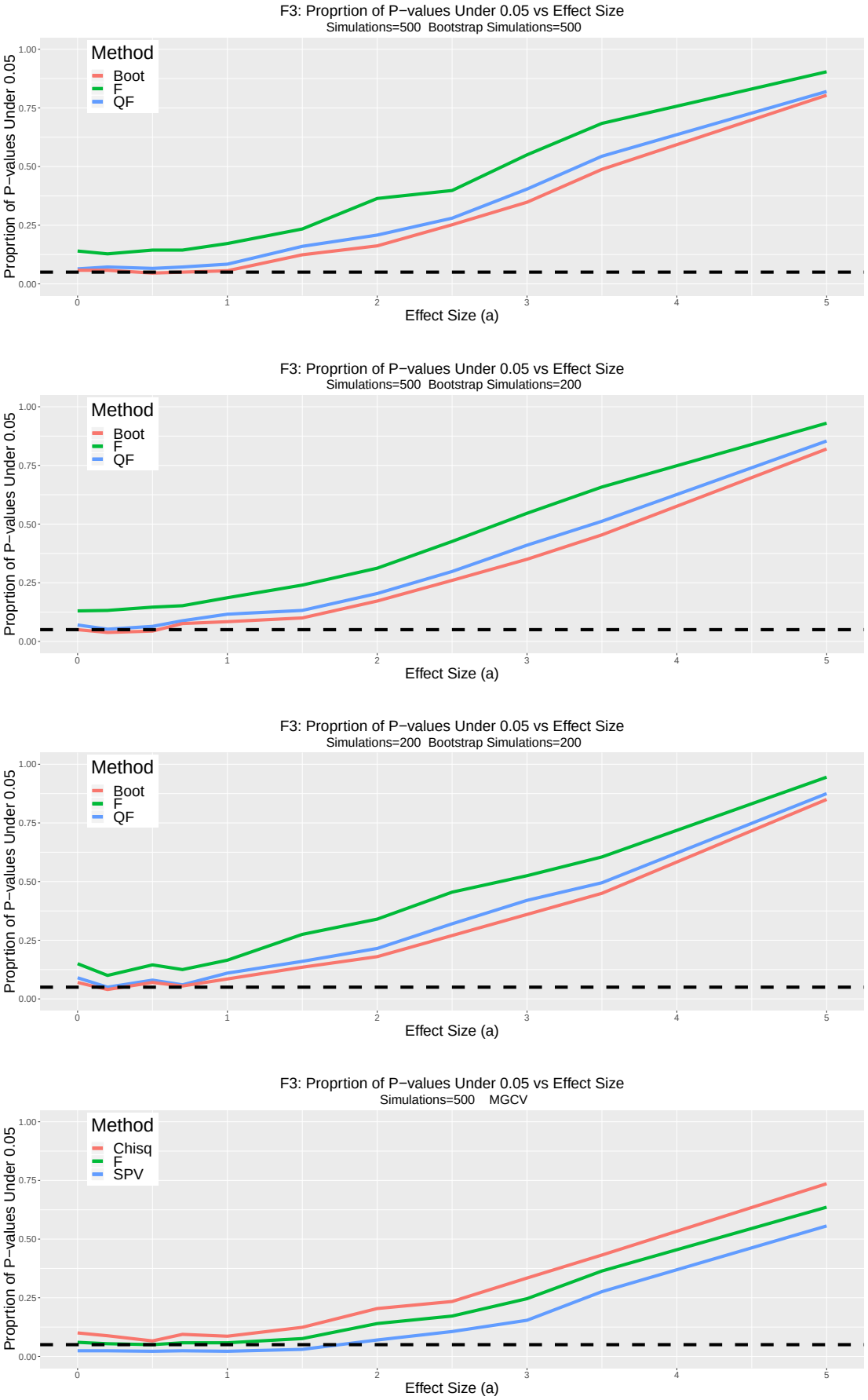


FIGURE 4.3: Plots showing the proportion of p-values under 5 % as a function of effect size for F3 for the 3 scenarios and `mgcv`.

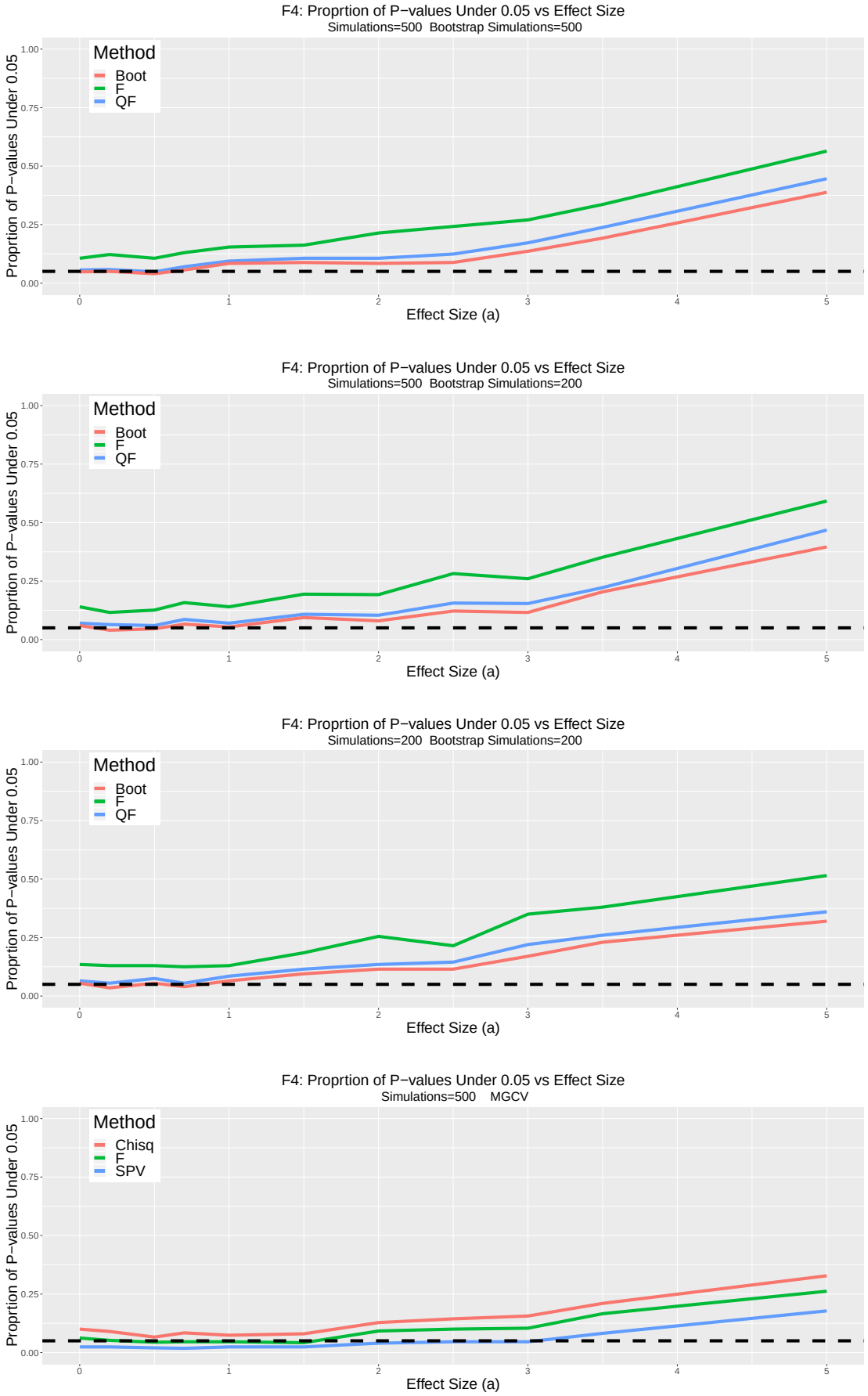


FIGURE 4.4: Plots showing the proportion of p-values under 5 % as a function of effect size for F4 for the 3 scenarios and `mgcv`.

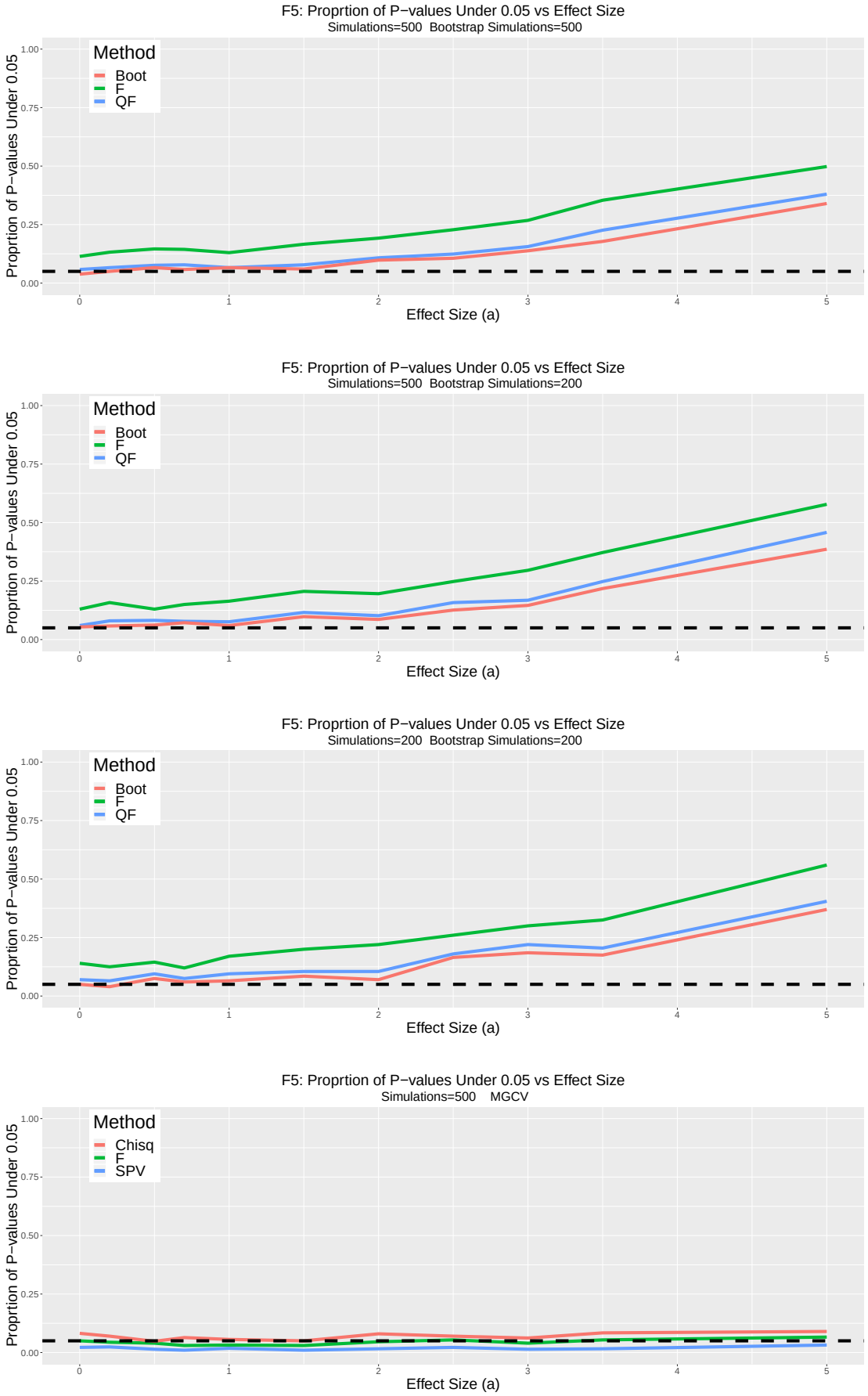


FIGURE 4.5: Plots showing the proportion of p-values under 5 % as a function of effect size for F5 for the 3 scenarios and `mgcv`.

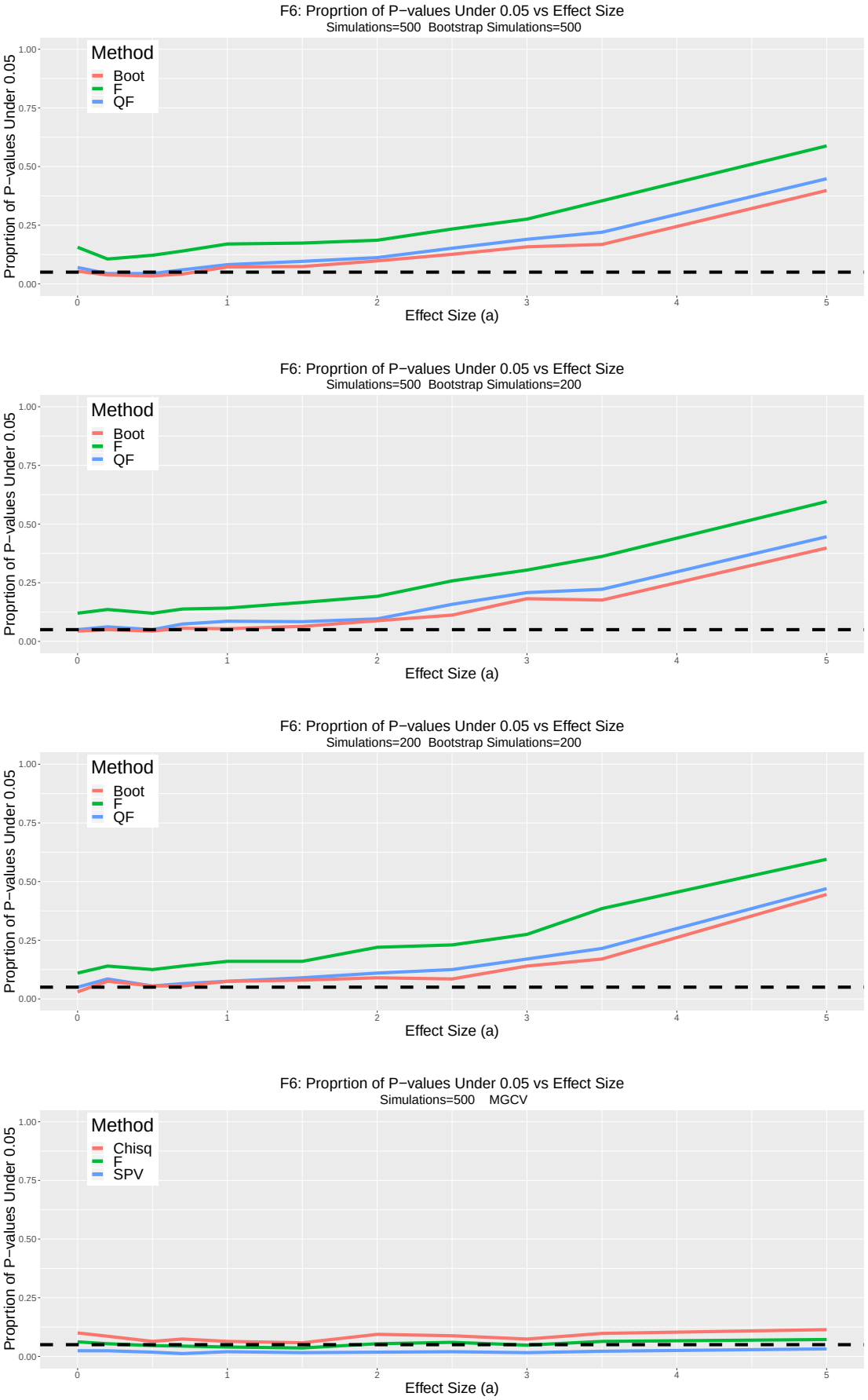
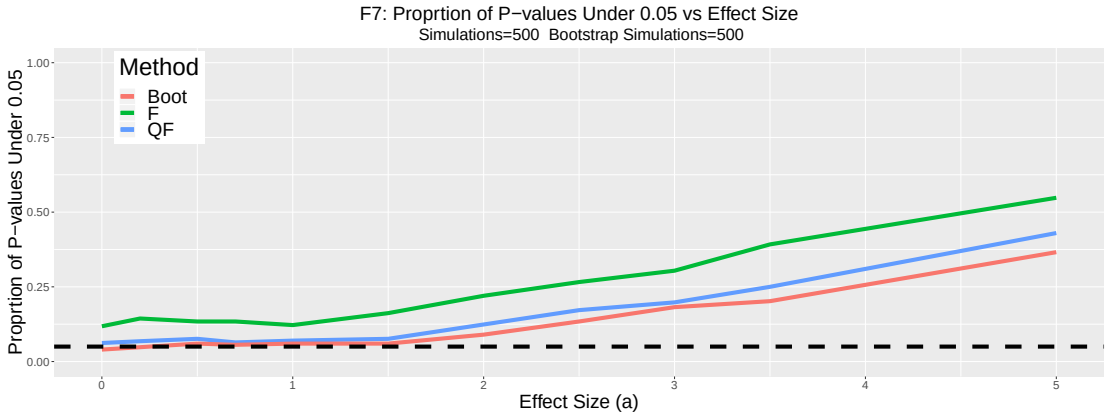
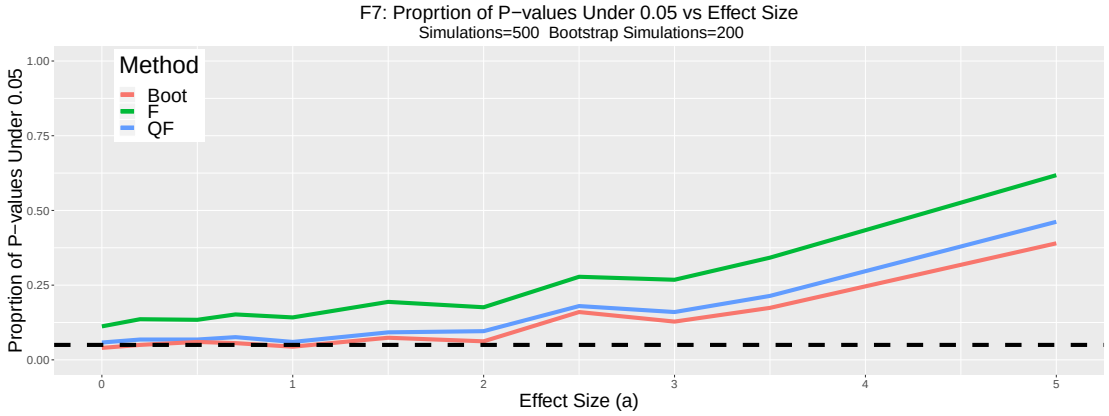
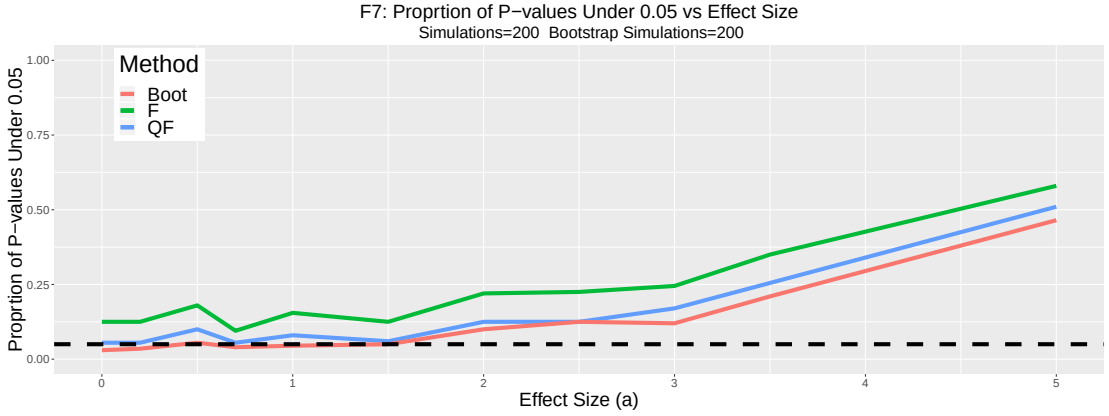


FIGURE 4.6: Plots showing the proportion of p-values under 5 % as a function of effect size for F6 for the 3 scenarios and `mgcv`.







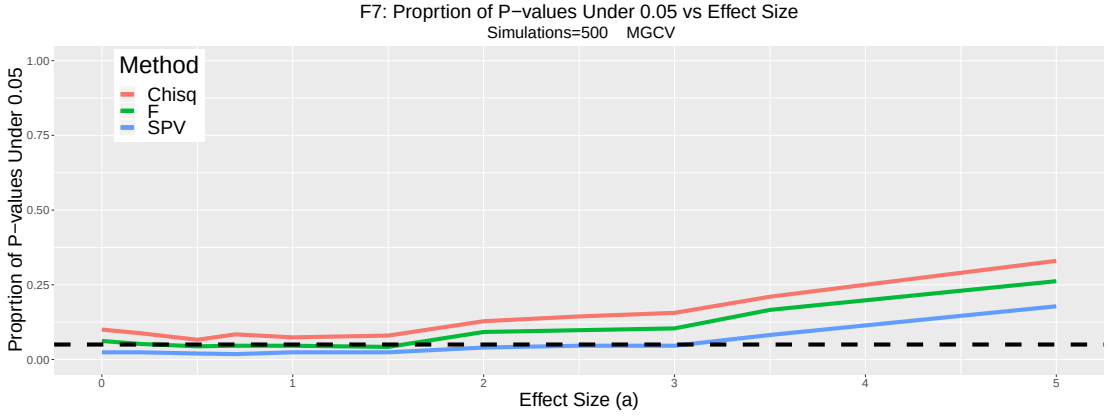
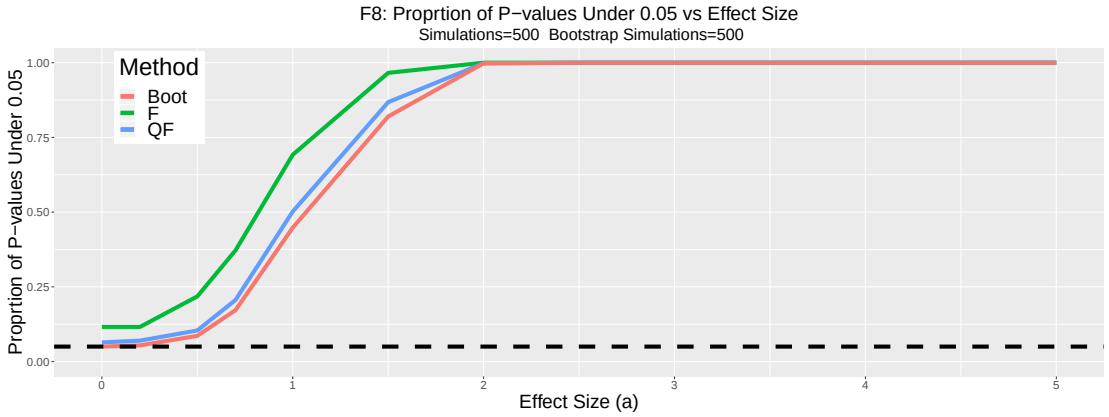
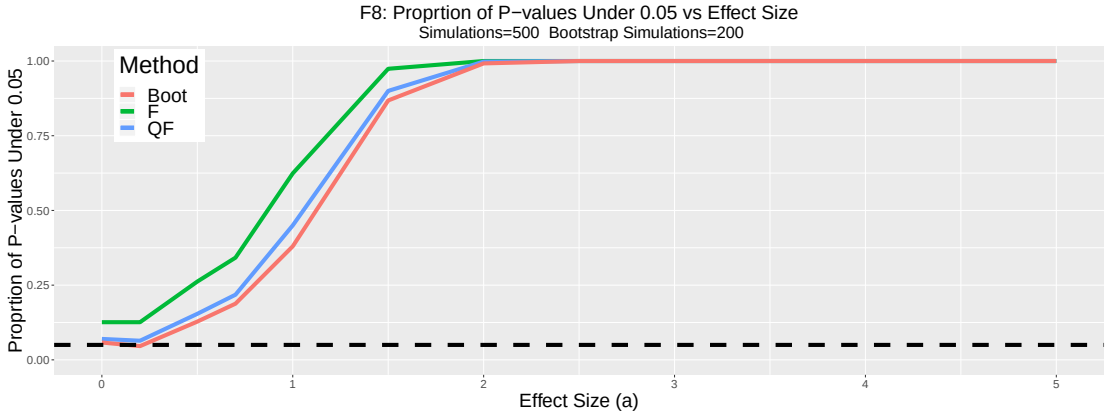
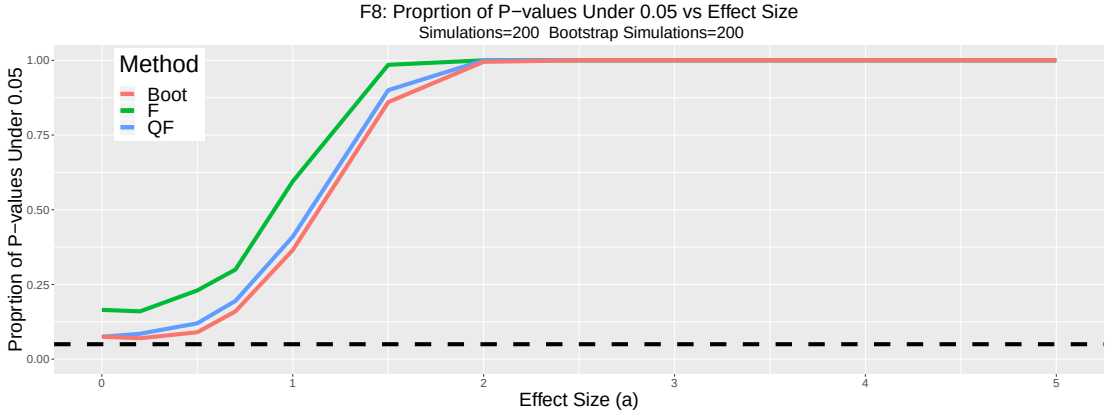


FIGURE 4.7: Plots showing the proportion of p-values under 5 % as a function of effect size for F7 for the 3 scenarios and `mgcv`.







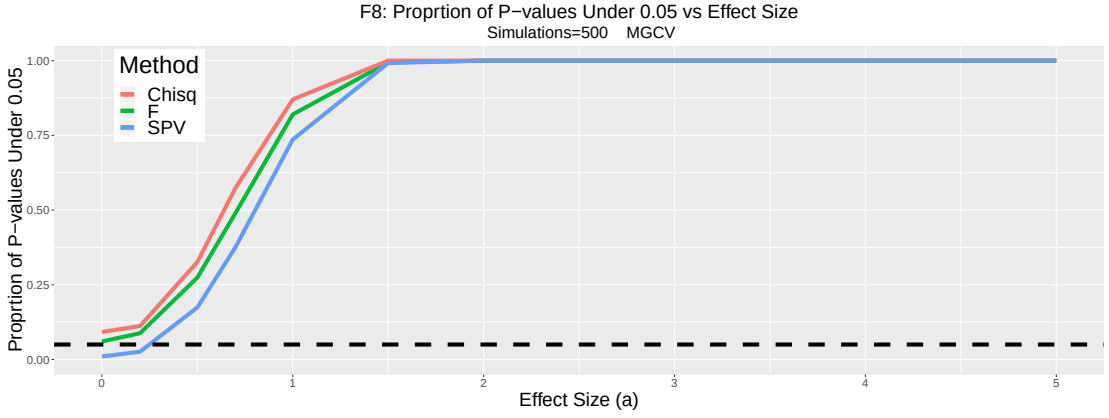


FIGURE 4.8: Plots showing the proportion of p-values under 5 % as a function of effect size for F8 for the 3 scenarios and `mgcv`.

4.4 Summary

Chapter 4 set out to examine the performance of two methods of ANOVA, the simple F-test and the F-test involving quadratic forms, implemented within the `sm` package. These methods were compared to a parametric bootstrap involving a simple F-test also within the `sm` package. A comparison was also made between these methods and the three methods embedded in `mgcv` (F-test, Chisq, and SPV). The quadratic forms method was shown to be well calibrated with more appropriate size compared to the simple F-test and superior power compared to the parametric bootstrap. Furthermore, the computational cost for the quadratic forms method was much lower than the parametric bootstrap. The `sm` methods generally outperformed the `mgcv` methods, specifically in terms of power.

Albeit the computational costs in using `sm` are greater than those of `mgcv`, the added control in implementing smoothing parameters will be essential when fitting an additive mixed model with interactions, and subsequently fitting a derivative additive mixed model, to the River Run data. Chapter 5 goes on to initially fit various additive models with no interactions to obtain a sense of the optimal degrees of freedom for each main effect. Various criterion scores will be implemented to see which of these scores performs best for the River Run data. Once a preferred criterion score has been identified, this score will be used to find the optimal degree of smoothing for each main effect. Interactions and subsequently random effects will be systematically implemented with optimal smoothing for the main effects and interactions to produce a final additive mixed model with interaction for the River Run data.

Chapter 5

Fitting a Model to the River Run Data

The River Run data includes many variables which may be potential drivers of the dissolved oxygen level. Furthermore, there may exist significant interactions between certain explanatory covariates' influence on the dissolved oxygen. Aside from presenting the raw data, this chapter will fit a simple P-spline model to describe the relationships between the dissolved oxygen and all explanatory covariates using a default degrees of freedom of 6 for each term. This initial model will contain no interaction terms and no random effects term. Subsequently, the various methods for smoothing parameter selection outlined in Chapter 3 will be applied to each explanatory covariate in the simple P-spline model. This will set a baseline for optimizing smoothing parameters for more complex P-spline models which will contain interaction terms and/or random effects. Models will be fitted with the corresponding optimal smoothing parameters and presented.

5.1 Fitting Simple and Complex Additive Models

The River Clyde is a complex system where the dissolved oxygen levels can be influenced by many natural phenomena and the activity of a substantial population of people along the river. It is reasonable to assume a representative statistical model should include several explanatory variables. It is also reasonable to assume there may exist significant interactions among particular variables. There is interest in seeing how some of these main effects on the dissolved oxygen may have changed as time passes and as we move down the river. Not only will bivariate interaction terms of each covariate with Year and each covariate with Station be implemented, but trivariate interaction terms of

each covariate with both Year and Station will be included. As complex as the model currently sounds, we must also consider other factors which apply on a particular day of sampling. A random effect component corresponding to sampling date may therefore also be needed to account for some of the variability. All additive models in this chapter will be fitted using `sm`, an R function to fit P-spline additive models, with a penalty order of 2 for all terms.

5.1.1 Simple Initial Additive Model

Before fine-tuning a more complex additive model by optimizing the smoothing parameters, it is worth constructing an initial additive model composed simply of main effects and seeing the relation each main effect has with the dissolved oxygen. This model will have the form

$$\begin{aligned} DO_i = & f_1(Day\ of\ Year_i) + f_2(Station_i) + \\ & f_3(Year_i) + f_4(Temperature_i) + \\ & f_5(Salinity_i) + f_6(Tide_i) + \\ & f_7(Spring_i) + f_8(RiverFlow_i) + \epsilon_i \end{aligned} \tag{5.1}$$

for $i = 1, ..., n$ where n represents the number of data points and ϵ represents the random error. The periodic natures of Day of Year, Spring, and Tide oblige us to implement periods of 365, 14.8 Haas (1977), and 200 respectively. This initial model is fitted with a default of 6 degrees of freedom for each of the main effects.

Figure 5.1 depicts all these main effects on the dissolved oxygen. Year (panel c) and Temperature (panel d) display patterns similar to the plots in Figure 2.7, albeit Temperature does display an unexpected upturn at the higher degrees Celsius. This slight increase in the Temperature main effect at approximately 17 degrees Celsius is worth noting. When consulting Dr. Alan Hills and Dr Ted Schlicke of SEPA about this behavior, I was informed this increase in dissolved oxygen during warmer weather could be attributed to photosynthesis of algae on sunny days. Station (panel b), Tide (panel f), Spring (panel g), and River Flow (panel h) seem plausible when looking back at Figure 2.7.

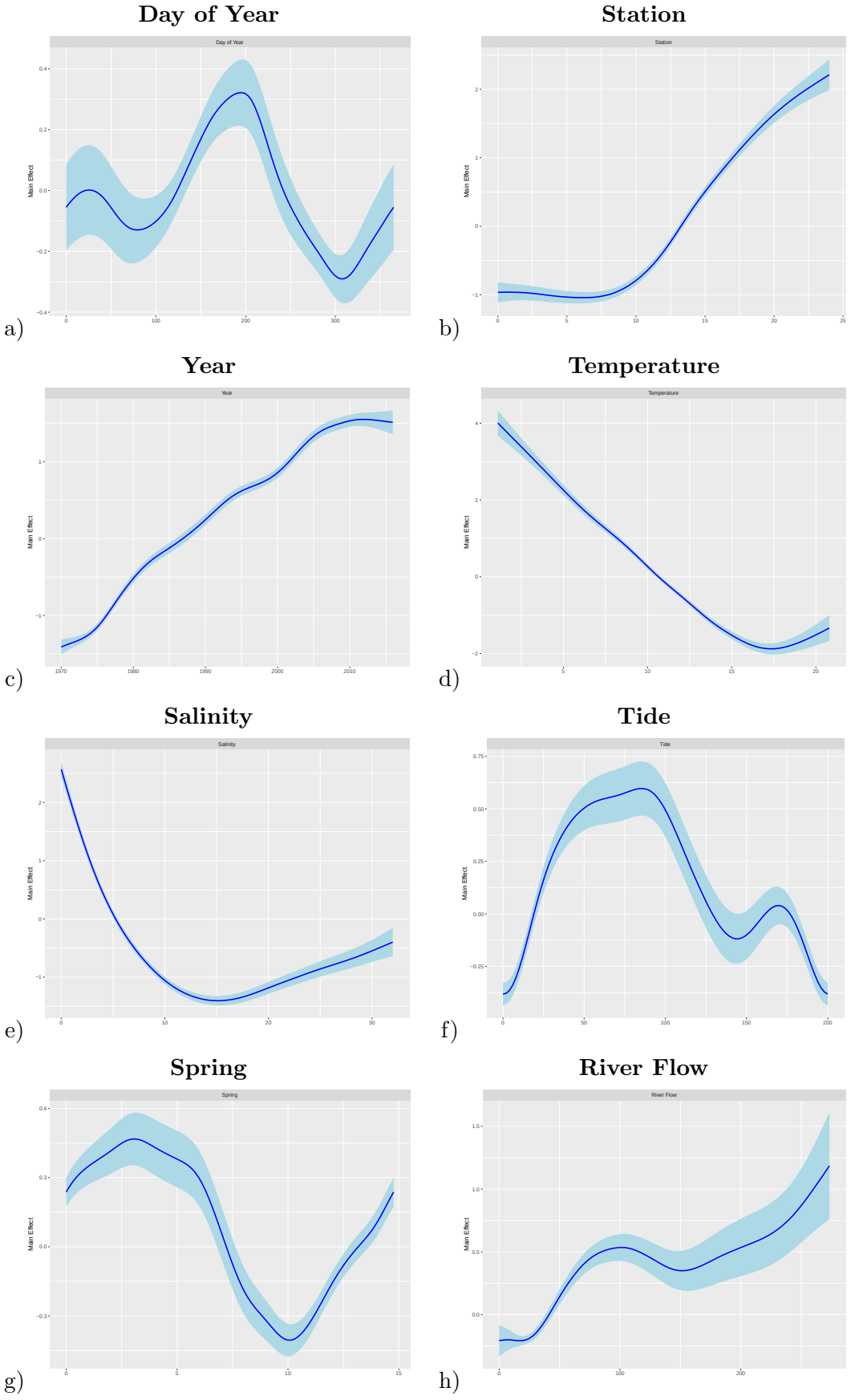


FIGURE 5.1: Plots of Main Effects a) Day of Year, b) Station, c) Year, d) Temperature, e) Salinity, f) Tide, g) Spring, and h) River Flow on DO for simple additive model and all degrees of freedom set to 6.

The main effects which are noticeably different from Figure 2.7 are Day of Year (panel a) and Salinity (panel e). Multicollinearity may be playing a role in why the marginal smooths are different than the smooths of the fitted values. Day of Year (Figure 5.1a) in conjunction with Temperature (Figure 5.1d) may warrant a closer look. The peak for the main effect of Day of Year coincides with the trough of the raw data scatter plot. I am assuming the range of the Temperature main effect adjustment, approximately [-2,4], being greater than the range of the Day of Year main effect adjustment, approximately [-0.3,0.3], would explain the peak of the Day of Year main effect to be in the summer months. In other words, the bulk of the adjustment to the dissolved oxygen on sunny days is made by the Temperature main effect and the Day of Year main effect subsequently compensates.

5.1.2 Simple Additive Model with Adjusted Smoothing Parameters

Having fitted a simple additive P-spline model in the previous section with a 6 degrees of freedom default for each main effect, the next step is to adjust the degrees of freedom for proper variance bias trade off. The methods outlined in Section 3.5 offer multiple means of optimizing the amount of smoothing. Minimizing the GCV, AIC, AICc, and BIC scores will yield the optimal degrees of freedom for each term. Although these methods may perform well for small data sets, there is a concern that the large number of data points in the River Run data set, just under 8000 for depths less than 2 meters, may cause problems in finding those minima.

The approach to finding the optimal degrees of freedom for each main effect taken here is as follows:

1. Select the first main effect of interest.
2. Define the range of all degrees of freedom to be applied to the main effect of interest.
3. Apply the first degrees of freedom to the main effect of interest.
4. Set all other main effect degrees of freedom to 6.
5. Fit the model.
6. Calculate GCV, AIC, AICc, and BIC scores.
7. Apply the next degrees of freedom to the main effect of interest.
8. Repeat steps 5 through 7 until all degrees of freedom have been applied.

9. Select the next main effect of interest.

10. Repeat steps 2 through 9 until all main effects have been assigned an optimal degrees of freedom.

Initially an exclusively GCV approach was considered. The range of degrees of freedom for each of the main effects initially was [3,12]. When the minimum GCV score happened to fall on the boundary values of the degrees of freedom for any one main effect, the range was expanded accordingly to achieve the desired local minimum. It proved neccessary to keep increasing the upper limit of degrees of freedom for several of the main effects because the maximum degrees of freedom was consistently being chosen as optimal. At this point the decision was made to include AIC, AICc, and BIC as alternative methods of smoothing parameter selection. Figure 5.2 depicts the trend of the four scores as the degrees of freedom increase across all main effects. Not all main effects run through the full range of degrees of freedom ([2,36]). Larger degrees of freedom values for the Spring and Station main effects caused singularity errors upon constructing the models. This is due to the degrees of freedom exceeding the 17 unique values of Spring and the 13 unique values of Station.

Additive models have the attractive characteristic of flexibility. However, too much flexibility yields an unrealistic model. One could expect certain main effects, such as Temperature and Salinity, in the simple additive model to have low degrees of freedom due to known physical properties of dissolved oxygen. One could also expect somewhat higher degrees of freedom for periodic main effects, such as Day of Year, Tide, or Spring. Year and Station have the potential for even higher degrees of freedom because all sorts of different factors can be in place at different times and locations. For these reasons the inability of GCV, AIC, and AICc to achieve minimums for a vast majority of these main effects over the degrees of freedom interval seems problematic. It seems unrealistic that some of these main effects would have optimal degrees of freedom greater than 36. Alternatively, the BIC score performed better than the rest as it managed to achieve a local minimum for every main effect and did so at lower degrees of freedom than did the other methods. This is due to the $log(n)$ term making BIC more suitable for data where n is large.

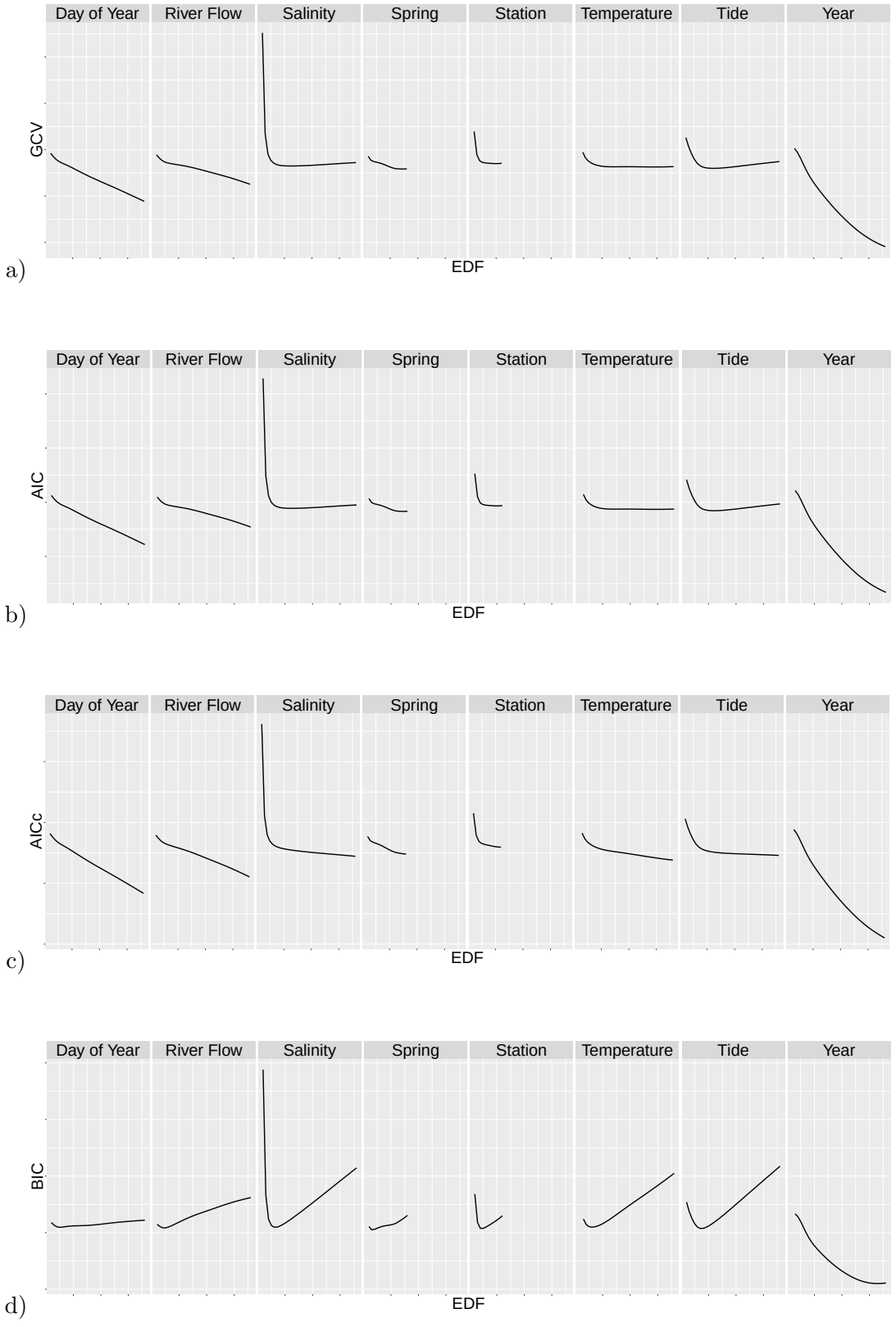


FIGURE 5.2: Plots depicting a) GCV, b) AIC, c) AICc, and d) BIC scores vs the estimated degrees of freedom (EDF) across each main effect.

The approach to finding the optimal degrees of freedom for each main effect taken here is as follows:

1. Select the first main effect of interest.
2. Define the range of all degrees of freedom to be applied to the main effect of interest.
3. Apply the first degrees of freedom to the main effect of interest.
4. Set all other not yet optimized main effect degrees of freedom to 6.
5. Fit the model.
6. Calculate the BIC score.
7. Apply the next degrees of freedom to the main effect of interest.
8. Repeat steps 5 through 7 until all degrees of freedom have been applied.
9. Set the degrees of freedom of main effect of interest to optimal value as chosen by the lowest BIC score.
10. Select the next main effect of interest.
11. Repeat steps 2 through 9 until all main effects have been assigned an optimal degrees of freedom.

The sequence the main effects degrees of freedom were optimized was

1. Day of Year
2. Temperature
3. Salinity
4. Spring
5. Tide
6. River Flow
7. Station
8. Year

The order in which the degrees of freedom are optimized may have an effect on the final optimized solution. This may be more evident when there are covariates combined in interaction terms. In the interest of consistency, the order chosen above is the order which will be used when fitting an additive mixed model with interactions later in this chapter where the justification for this order will also be presented.

Now that BIC has been identified as the smoothing parameter selection method of choice and a baseline of degrees of freedom for each term has been set, the task at hand is finding the optimal degrees of freedom for each term of model 5.1 and of a comparable model with random effects ν

$$\begin{aligned} DO_{ij} =& f_1(Day\ of\ Year_{ij}) + f_2(Station_{ij}) + \\ & f_3(Year_{ij}) + f_4(Temperature_{ij}) + \\ & f_5(Salinity_{ij}) + f_6(Tide_{ij}) + \\ & f_7(Spring_{ij}) + f_8(RiverFlow_{ij}) + \nu_j + \epsilon_{ij} \end{aligned} \tag{5.2}$$

where j is an index on the sampling date. Table 5.1 shows the optimal degrees of freedom as chosen by the minimum BIC score for simple additive models without and with random effects respectively.

Main Effect	Degrees of Freedom-BIC (No Random Effects)	Degrees of Freedom-BIC (With Random Effects)
Day of Year	5	5
Temperature	6	6
Salinity	6	9
Spring	3	5
Tide	7	8
River Flow	5	7
Station	5	4
Year	33	46

TABLE 5.1: Table of main effects and corresponding degrees of freedom as chosen by BIC for simple additive model without and with random effects.

Figures 5.3 and 5.4 depict the main effects of models 5.1 and 5.2 respectively. With the exception of Day of Year, almost all main effects exhibit little change when a random effects component is added to the model. The change in Day of Year is however worth noting. It is expected the Day of Year main effect to have periodic behaviour. Both models 5.1 and 5.2 do exhibit this behaviour, but the latter has a Day of Year that is more plausible because it has a simpler sinusoidal form.

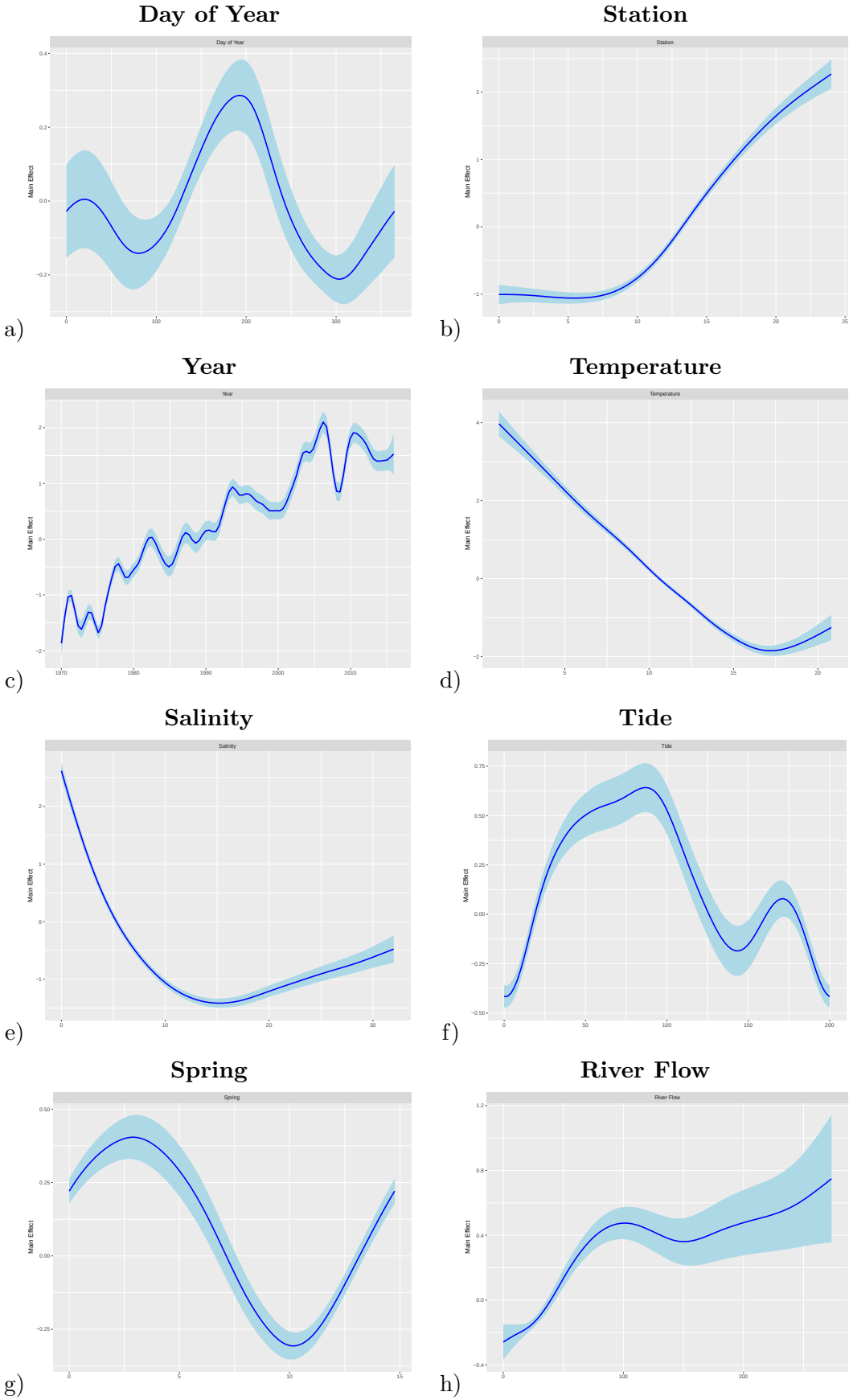


FIGURE 5.3: Plots of Main Effects a) Day of Year, b) Station, c) Year, d) Temperature, e) Salinity, f) Tide, g) Spring, and h) River Flow on DO for simple additive model using optimized degrees of freedom as chosen by BIC.

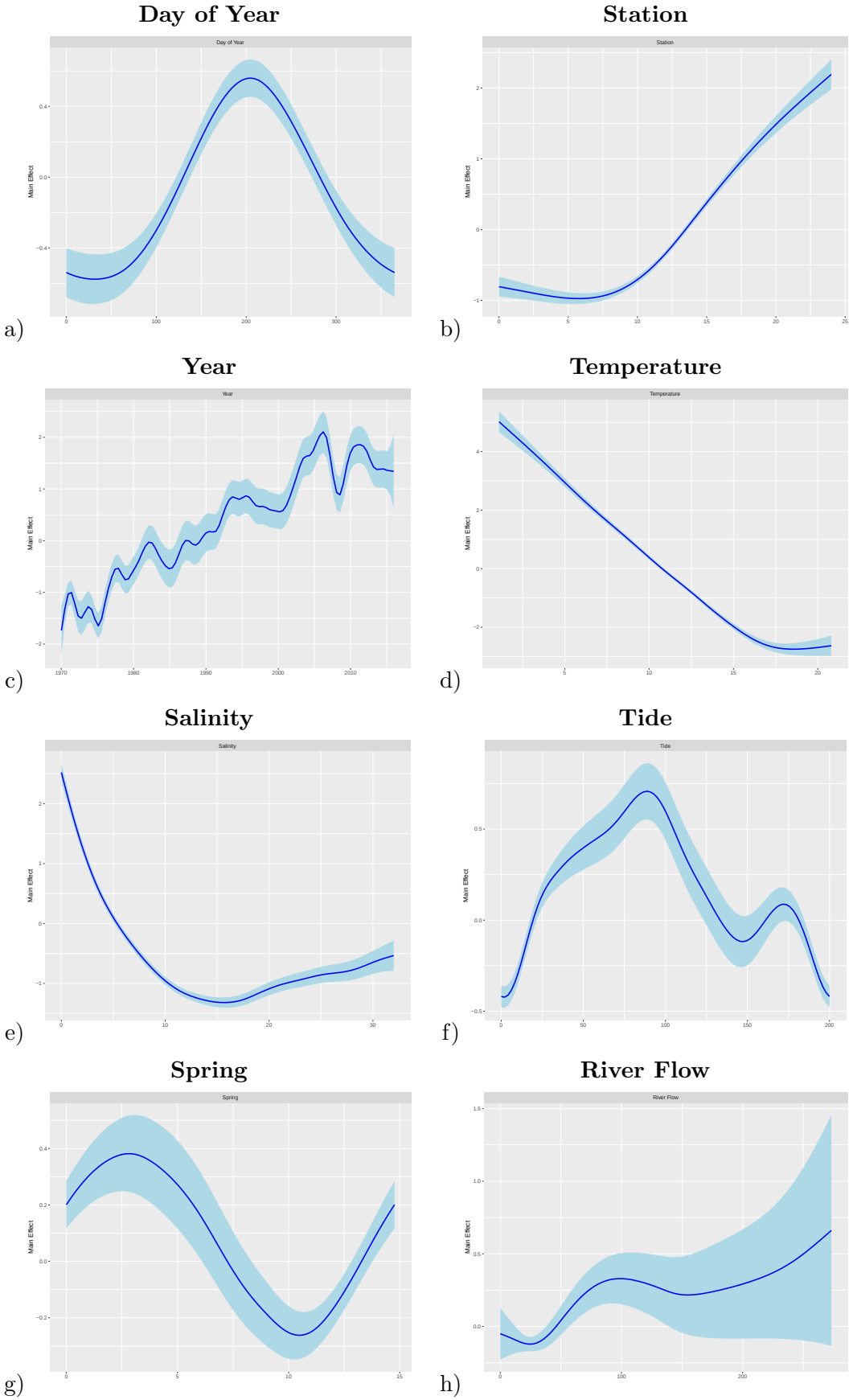


FIGURE 5.4: Plots of Main Effects a) Day of Year, b) Station, c) Year, d) Temperature, e) Salinity, f) Tide, g) Spring, and h) River Flow on DO for simple additive mixed model using optimized degrees of freedom as chosen by BIC.

5.1.3 Smoothing Parameter Selection for Bivariate Additive Models with Interactions

There is interest in seeing if the order covariates are smoothed influences the optimal degrees of freedom selected for additive models with interaction terms. A simulation study would help to bear out the orders importance. Consider the following bivariate functions with interactions included:

- **F1**: $y = 2cos(2\pi x_1) + 2x_2 + 3e^{x_1 x_2}$
- **F2**: $y = 2sin(2\pi x_1) + 2e^{x_2} + 5(x_1 x_2)^2$
- **F3**: $y = 2sin(2\pi x_1) + 2e^{x_2} + 5\sqrt{x_1 x_2}$
- **F4**: $y = 2sin(2\pi x_1) + 2e^{-x_2} + 5\sqrt{x_1 x_2}$
- **F5**: $y = 2sin(2\pi x_1) + 2e^{-x_2} + (sin(2\pi x_1)sin(2\pi x_2))$

These functions will generate data to which random normal noise can be added. For each simulation x_1 will have its degrees of freedom optimized before x_2 and then vice versa. All four degrees of freedom will be recorded before the the next simulation generates a new set of data from the same function. This is done for 500 simulations. The results can be plotted. The functions **F1** through **F5** are similar to the ones used in Chapter 4 so the interactions are known to be detectable. Standard deviations of 1 and 2 will be used in the simulations and compared.

The distributions of the optimized degrees of freedom for the main effects of functions **F1** through **F5** are depicted in Figures 5.5 and 5.6. The horizontal axis represents the degrees of freedom which are whole number values. Each whole number is centered at one grid line and extends one grid line in each direction. This means any vertical bar falling within one grid line of the whole number is within that value of degrees of freedom. Both x_1 and x_2 have a *First* and *Second* description. If the order of optimization is unimportant, the *First* and *Second* for x_1 will be similar and the *First* and *Second* for x_2 will be similar. This generally seems to be the case for all 5 function across both standard deviations, with only modest differences in the degrees of freedom identified. This offers some reassurance that the order in which variables are considered does not make a crucial difference.

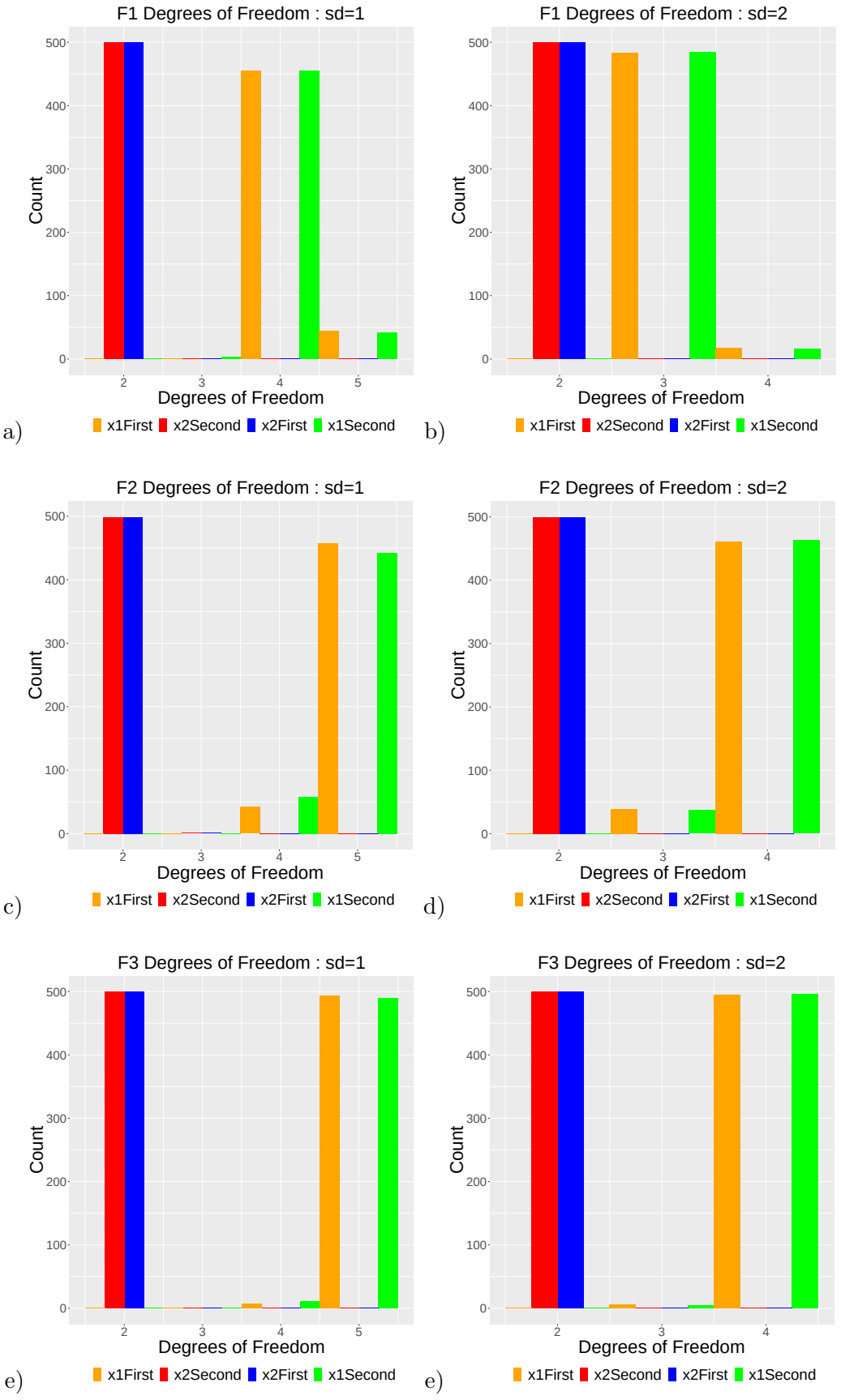


FIGURE 5.5: Plots of distributions of selected degrees of freedom for functions F1 through F3. The left column represents standard deviation 1 and the right column represents standard deviation 2

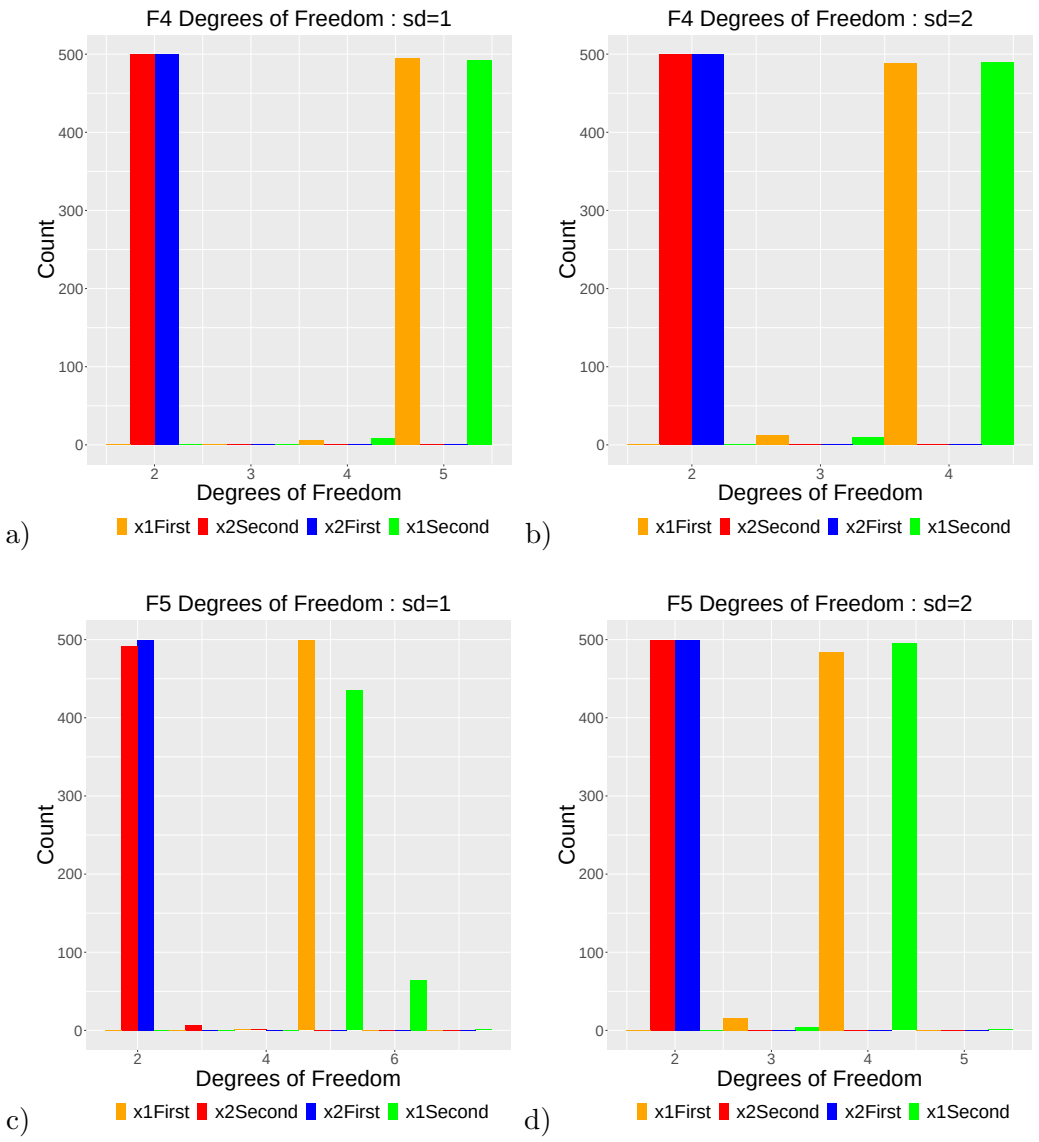


FIGURE 5.6: Plots of distributions of selected degrees of freedom for functions F4 and F5. The left column represents standard deviation 1 and the right column represents standard deviation 2.

5.1.4 Complex Additive Mixed Model with Adjusted Smoothing Parameters

The River Clyde is a complex system which we would be hard pressed to represent with an additive model with no of interactions. It is reasonable to assume that the six effects of Day of Year, Temperature, Salinity, Tide, Spring, and River Flow all change to some degree at different sampling stations. Furthermore, these same six effects may also have altered in complex ways over the years. Interaction terms of the six effects with Station, the six effects with Year, and Station with Year would make the appropriate adjustments to compensate for these subtle changes. Including these bivariate interaction terms may

still not be enough to properly represent such a complex system. Station in conjunction with Year may also play a role in degree of necessary adjustment to be made in addition to the main effects and bivariate interactions. Trivariate interaction terms of each of the six effects with both Station and Year will be included.

It is reasonable to believe there are certain factors in place during a particular day of sample collection which may cause the dissolved oxygen levels to collectively be higher or lower than some other day. These factors may not be measurable or may have been overlooked altogether. Thus, the sampling regime of multiple samples being taken on the same day justifies the inclusion of a random effects component. This inclusion of the random effects components gives rise to the final additive mixed model with interactions.

The complex additive mixed model to be fitted is composed of main effects, bivariate interaction terms, trivariate interaction terms, and a random effects term. The notation depicting such a model can become extensive when there are many explanatory variables. In the interest of brevity, an asterisk (*) will be used to denote all possible main effects and multivariate interactions are included. An example would be

$$y = f_1(x_1) * f_2(x_2) * f_3(x_3) \tag{5.3}$$

which can be interpreted as

$$\begin{aligned} y = & f_1(x_1) + f_2(x_2) + f_3(x_3) \\ & + f_{1,2}(x_1, x_2) + f_{1,3}(x_1, x_3) + f_{2,3}(x_2, x_3) \\ & + f_{1,2,3}(x_1, x_2, x_3) \end{aligned} \tag{5.4}$$

The notation in equation 5.3 now allows for the expresion of the complex additive mixed model

$$\begin{aligned} DO_{ij} = & f_1(Day\ of\ Year_{ij}) * f_2(Station_{ij}) * f_3(Year_{ij}) \\ & + f_4(Temperature_{ij}) * f_2(Station_{ij}) * f_3(Year_{ij}) \\ & + f_5(Salinity_{ij}) * f_2(Station_{ij}) * f_3(Year_{ij}) \\ & + f_6(Tide_{ij}) * f_2(Station_{ij}) * f_3(Year_{ij}) \\ & + f_7(Spring_{ij}) * f_2(Station_{ij}) * f_3(Year_{ij}) \\ & + f_8(RiverFlow_{ij}) * f_2(Station_{ij}) * f_3(Year_{ij}) + \nu_j + \epsilon_{ij} \end{aligned} \tag{5.5}$$

where j is an index on the sampling date.

Once again, since BIC has been identified as the smoothing parameter selection method of choice and a baseline of degrees of freedom for each term has been set, the task at hand is finding the optimal degrees of freedom for each term of a more complex model. This will be an additive mixed model with interaction terms.

Section 5.1.3 showed permuting the order the main effects had little material effect on the optimal degrees of freedom of both covariates for bivariate additive models with interactions. The additive mixed model with interactions for the River Run data is much more complex. Therefore, it is logical to see if different permutations of the optimizing order give substantially different degrees of freedom for the main effects. Since there are 8 main effects, going through all permutations would be very time consuming. Instead, 4 different permutations are considered. The original order is the same as what was used in fitting the simple additive model with no interactions. This permutation primarily sought to optimize Station and Year last because these were the two main effects which would be involved in the interactions with the other covariates. The first 6 covariates were placed in an order where there exists or may exist some correlation between sequentially adjacent covariates. The second permutation is reverse order of permutation 1. Permutation 3 considered the degrees of freedom of the simple additive mixed model given in Table 5.1 and put them in descending order from Year with the highest degree of freedom (46) to Station with the lowest degree of freedom (4). Permutation 4 was the reverse order of permutation 3.

Table 5.2 shows the optimal degrees of freedom as chosen by minimum BIC score for the complex additive mixed model for all permutations. It is apparent, for permutation 1, all but one term (Spring for no random effects) have degrees of freedom lower than those of the simple models shown in Table 5.1, with Year degrees of freedom being substantially lower. All optimized degrees of freedom showed little change throughout the 4 permutations, with Salinity showing the biggest change from 6 to 9. Permutations 2, 3, and 4 all found optimal degrees of freedom greater than or equal to those found in permutation 1. This result shows the order the terms are smoothed does influence the final optimal degrees of freedom but the effect is small.

5.1.4.1 Main Effects

The main effects for model 5.5 are depicted in Figure 5.7. The general decrease in the degrees of freedom of the main effects of this model is reflected in the reduced flexibility of those main effects when interaction terms are included. The most dramatically changed main effects are Station, Year, and River Flow, which have become nearly linear. Tide has also changed to a more plausible simple sinusoidal form.

Permutation 1

Main Effect	Degrees of Freedom - BIC
Day of Year	4
Temperature	5
Salinity	6
Spring	4
Tide	4
River Flow	2
Station	2
Year	2

Permutation 2

Main Effect	Degrees of Freedom - BIC
Year	2
Station	2
River Flow	2
Tide	4
Spring	4
Salinity	9
Temperature	6
Day of Year	5

Permutation 3

Main Effect	Degrees of Freedom - BIC
Year	2
Salinity	7
Tide	4
River Flow	2
Temperature	6
Spring	4
Day of Year	5
Station	4

Permutation 4

Main Effect	Degrees of Freedom - BIC
Station	2
Day of Year	5
Spring	4
Temperature	6
River Flow	2
Tide	4
Salinity	8
Year	2

TABLE 5.2: Table of main effects and corresponding degrees of freedom as chosen by BIC for complex additive mixed model with interactions for all 4 permutations.

Figure 5.8 depicts the same main effects as Figure 5.7 but on a broader scale. The distances the points are from the main effect depicted in blue represent the partial residuals. The partial residuals are the distance the raw data are from the predicted value of a particular main effect when all other main effects have been removed. These plots allow one to see in detail the role each main effect is playing on the model as a

whole. The main effects of the additive mixed model with interactions can be described as follows:

- Day of Year - This main effect takes on a sinusoidal pattern with minimums of approximately -0.2 in the winter months and a maximum of 0.2 in the summer months.
- Station - This main effect is relatively positive linear with a minimum of approximately -1.5 at Station 0 and a maximum of approximately 1.8 at Station 24.
- Year - This main effect is relatively positive linear with a minimum of approximately -1.5 at Year 1970 and a maximum of approximately 2.3 at Year 2016.
- Temperature - This main effect is relatively negative linear with a maximum of approximately 4.1 at Temperature 1 degree Celsius and a minimum of approximately -1.9 at Temperature 17 degrees Celsius. There is a slight increase beginning at 17 and ending at 21 degrees Celsius.
- Salinity - This main effect takes on a parabolic pattern with a maximum of 2.0 at Salinity 0 and a minimum of -1.2 at Salinity 18. There is a slight increase from Salinity 18 to Salinity 32.
- Tide - This effect takes on a sinusoidal pattern with a maximum of 0.6 just before Tide 100 (high diurnal tide) and a minimum of -0.2 at Tide 0 and 200 (low diurnal tide).
- Spring - This effect takes on a sinusoidal pattern with a maximum of 0.4 at Spring 3 (just before spring tide) and a minimum of -0.2 at Spring 11 (just after spring tide).
- River Flow - This main effect is relatively positive linear with a minimum of approximately -0.3 at River Flow 0 and a maximum of approximately 1.5 at River Flow 280.

These are the roles each predictor is playing as a main effect on this additive mixed model with interactions.

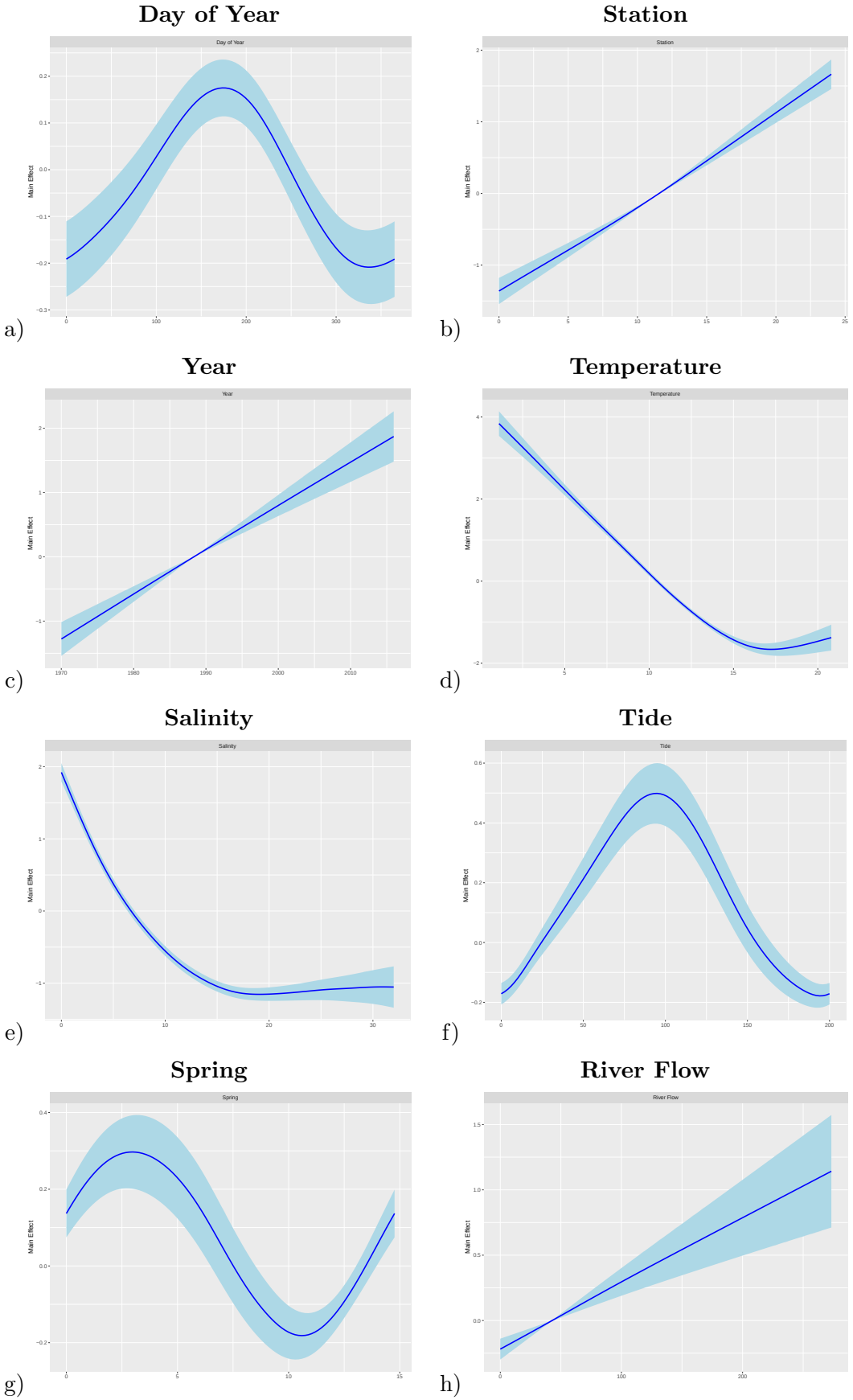


FIGURE 5.7: Plots of Main Effects a) Day of Year, b) Station, c) Year, d) Temperature, e) Salinity, f) Tide, Spring, and g) River Flow on DO for complex additive mixed model using optimized degrees of freedom as chosen by BIC.

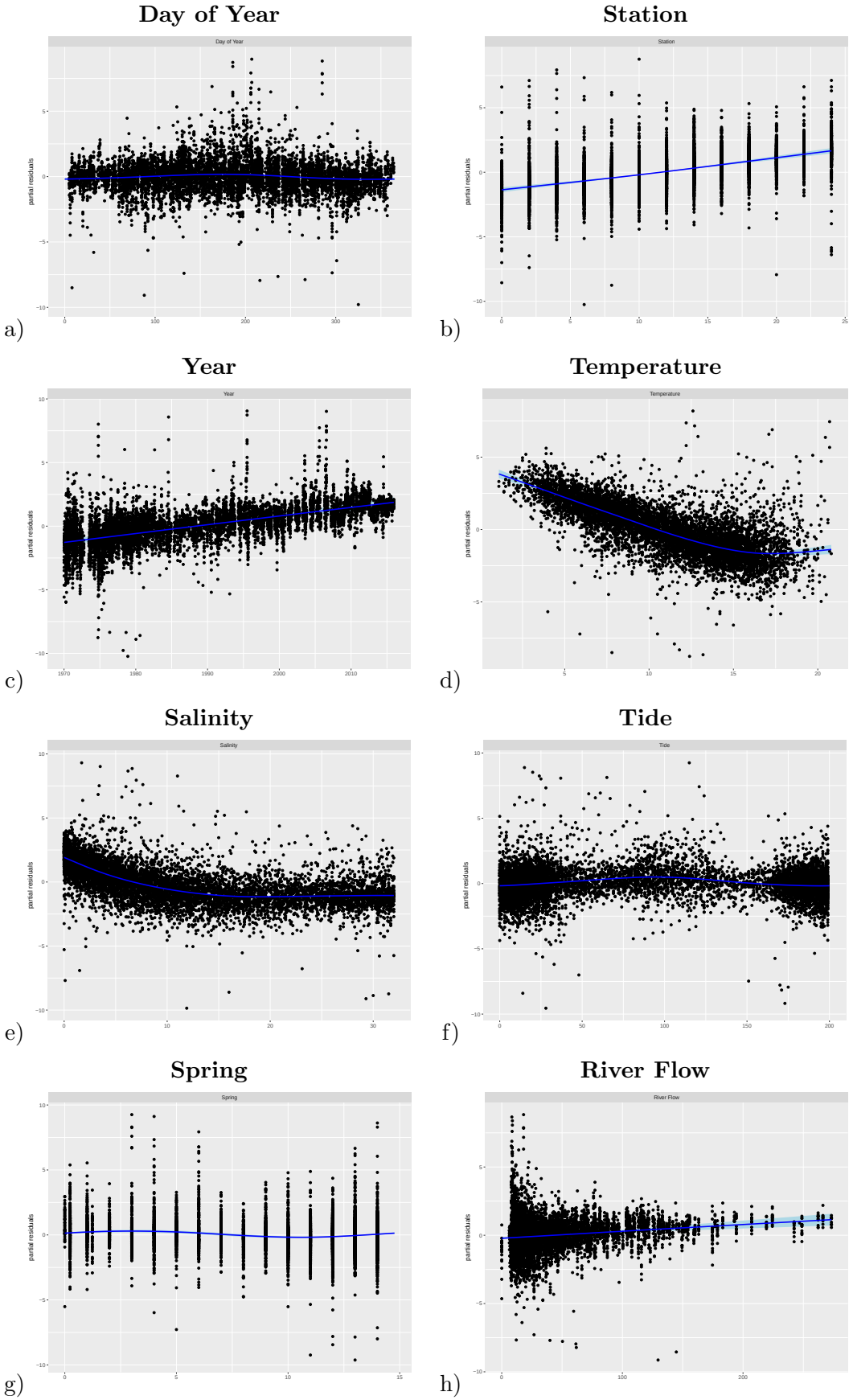


FIGURE 5.8: Partial residual plots of Main Effects a) Day of Year, b) Station, c) Year, d) Temperature, e) Salinity, f) Tide, Spring, and g) River Flow on DO for complex additive mixed model.

5.1.4.2 Bivariate Interaction Terms

For clarity, the phrase "lower terms" will be defined as the main effects associated with a bivariate interaction term or the main effects and bivariate interaction terms associated with a trivariate interaction term. The plotting utility within `sm` allows the depiction of the interactions with and without lower terms included. This section will utilize plots of interactions with lower terms included, postponing plots depicting interactions without lower terms included until Section 5.1.4.3. These interaction are represented as surfaces in perspective and contour plots in Figures 5.9-5.13. The overall effect on the dissolved oxygen is easy to determine. What may be difficult to discern in some of these surfaces is the degree of adjustment contributed by the interaction. However, the degrees of these adjustments are represented by varying colours in the perspective plots on the left and coloured contour lines in contour plots on the right. The perspective plots use position along the DO (dissolved oxygen) axis to signify the level of dissolved oxygen. The perspective plots also use the progression of dark blue, light blue, green, yellow, and red to signify a negative adjustment to a positive adjustment with green signifying no adjustment. The contour plots offer a more refined depiction for detecting these adjustments. The numbering on the contours represents the amount of standard errors the surface is adjusted from no interaction effect, where maroon and blue represent a positive and negative adjustment respectively. The colouring beneath the contours represents the interaction adjustment added to the underlying main effects, where maroon and blue represent high and low levels of dissolved oxygen respectively. For this reason the higher level contours frequently do not coincide with the higher level of dissolved oxygen.

Upon inspection of Figures 5.9 - 5.13, it is clear the colouring under the contours in the contour plots corresponds to the height of the surface in the perspective plots. Also clear is the colours on the perspective plots coincide with the contours in the contour plots. The behavior of the bivariate interactions can be summarized as follows:

- Day of Year : Station (Figure 5.9a-b) — Negative adjustments of at least -10 standard errors are present near Station 12 during the summer months and near Station 24 during the winter months. Positive adjustments of at least +10 standard errors are present near Station 12 during the winter months and near Station 24 during the summer months. There is very strong evidence the Day of Year : Station interaction is significant and should be included in the model.
- Day of Year : Year (Figure 5.9c-d) — Negative adjustments of at least -3 standard errors are present circa at 1975 during the spring months and circa 2002 during the autumn months. Positive adjustments of at least +3 standard errors are present

during a time period centered at 1975 during the autumn months and during a time period centered at 2002 during the spring months. There is strong evidence the Day of Year : Year interaction is significant and should be included in the model.

- Temperature : Station (Figure 5.9e-f) — Negative adjustments of at least -10 standard errors are present near Station 9 at lower temperatures and near Station 24 at higher temperatures. Positive adjustments of at least +12 standard errors are present near Station 9 at lower temperatures and near Station 24 at higher temperatures. There is very strong evidence the Temperature : Station interaction is significant and should be included in the model.

- Temperature : Year (Figure 5.10a-b) — Negative adjustments of at least -3 standard errors are present circa 1980 at higher temperatures and circa 2000 at lower temperatures. Positive adjustments of at least +2 standard errors are present during a time period centered at 1980 lower temperatures and during a time period centered at 2000 higher temperatures. There is strong evidence the Temperature : Year interaction is significant and should be included in the model.

- Salinity : Station (Figure 5.10c-d) — Negative adjustments of at least -12 standard errors are present the ordered pairs (Station=0 , Salinity=20) and (Station=14 , Salinity=5) with a wide band of negative contours along the line connecting these minima. Positive adjustments of at least +10 standard errors are present outside of the negative adjustment band. There is very strong evidence the Salinity : Station interaction is significant and should be included in the model.

- Salinity : Year (Figure 5.10e-f) — Negative adjustments of at least -8 standard errors are present circa 1975 at higher levels of salinity and circa 2010 for lower levels of salinity. Positive adjustments of up to +12 standard errors are present in a banded form between the negative adjustments. There is very strong evidence the Salinity : Year interaction is significant and should be included in the model.

- Spring : Station (Figure 5.11a-b) — Negative adjustments of at least -7 standard errors are present during and just after maximum spring tide at Stations 0-12 and just after neap tide at Stations 16-24. Positive adjustments of at least +3 or up to +7 are present in banded form just after neap tide at Stations 0-12 and progressing Stations 16-24 during and just after maximum spring tide. There is very strong evidence the Spring : Station interaction is significant and should be included in the model.

- Spring : Year (Figure 5.11c-d) — Negative adjustments of at least -2 standard errors are present just after maximum spring tide circa 1978 and just after neap

tide circa 2007. Positive adjustments of at least +2 are present just after neap tide circa 1978 and during and just after maximum spring tide circa 2006. There is weak evidence the Spring : Year interaction is significant and should be included in the model.

- Tide : Station (Figure 5.11e-f) — Negative adjustments of at least -6 or up to -12 standard errors are present during high tide around Stations 2 and 24 and during low tide around Stations 12 and 14. Positive adjustments of at least +6 or up to +12 are present during high tide around Stations 12 and 14 and during low tide around Stations 2 and 24. There is very strong evidence the Tide : Station interaction is significant and should be included in the model.

- Tide : Year (Figure 5.12a-b) — Negative adjustments of at least -6 standard errors are present during high tide circa 2000 and during low tide circa 1970. Positive adjustments of at least +6 are present during high tide circa 1970 and just after low tide circa 2000. There is very strong evidence the Tide : Year interaction is significant and should be included in the model.

- River Flow : Station (Figure 5.12c-d) — Negative adjustments of at least -12 standard errors are present for River Flows of 0-100 at Stations 0-12 and for River Flows above 120 for Stations 14-24. Positive adjustments of at least +12 standard errors are present for River Flows above 120 at Stations 0-12 and for River Flows of 0-100 for Stations 14-24. There is very strong evidence the River Flow : Station interaction is significant and should be included in the model.

- River Flow : Year (Figure 5.12e-f) — Negative adjustments of at least -2 or up to -3 standard errors are present for River Flows of 0-120 from 1970 to 1990 and for River Flows above 140 from 1993 to 2016. Positive adjustments of at least +3 standard errors are present for River Flows above 140 from 1970 to 1990 and for River Flows of 0-120 from 1993 to 2016. There is strong evidence the River Flow : Year interaction is significant and should be included in the model.

- Year : Station (Figure 5.13a-b) — Negative adjustments of at least -8 or up to -14 standard errors are present for the years 1970-1993 at Stations 0-12 and for the Years 1993-2016 at Stations 14-24. Positive adjustments of at least +8 or up to +12 standard errors are present for the Years 1994-2016 at Stations 0-12 and for the Years 1970-1993 for Stations 14-24. There is very strong evidence the Year : Station interaction is significant and should be included in the model.

The summary above shows, with the exception of Spring : Year, all bivariate interactions exhibit strong evidence of being significant and should be included in the River Run additive mixed model.

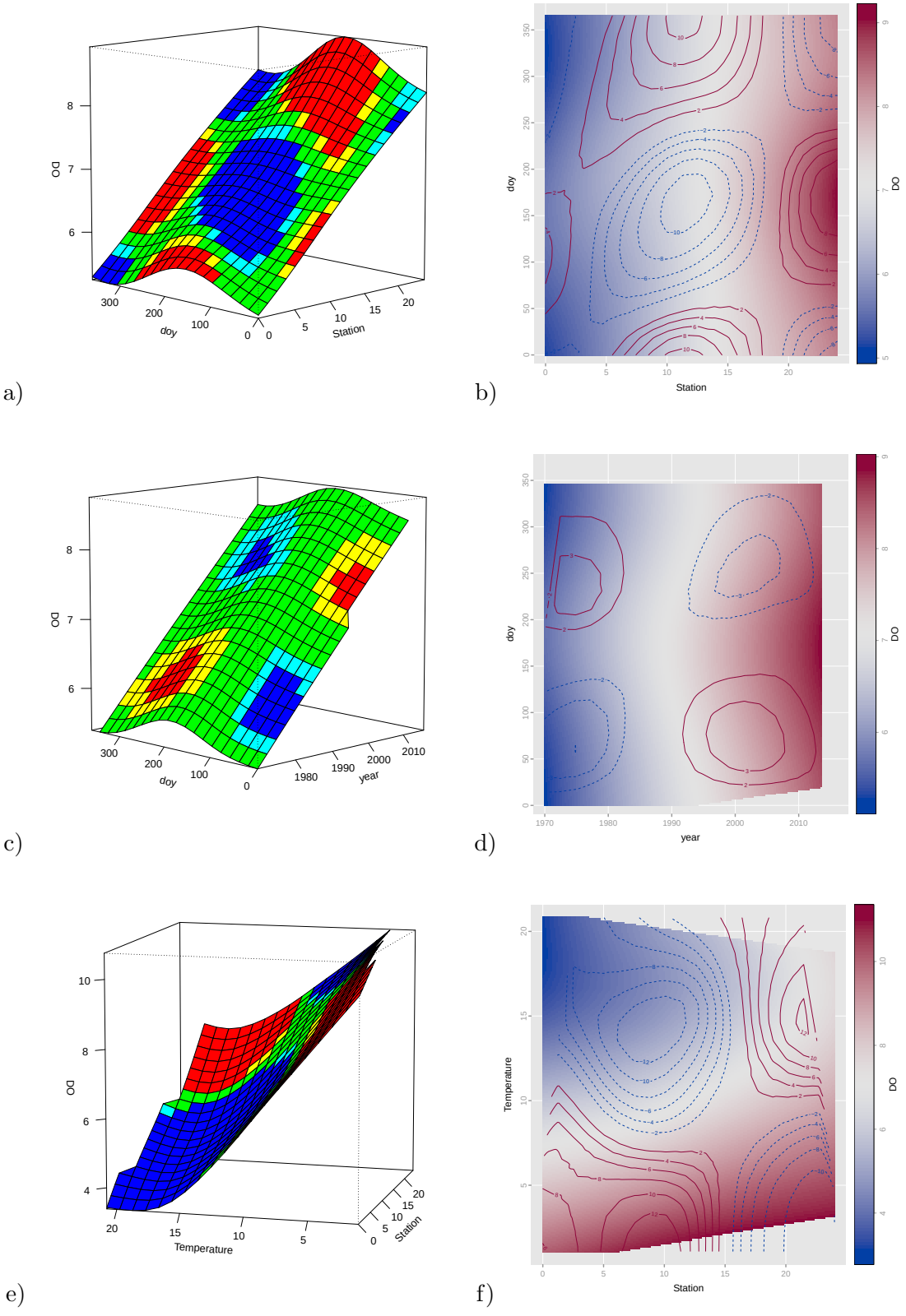
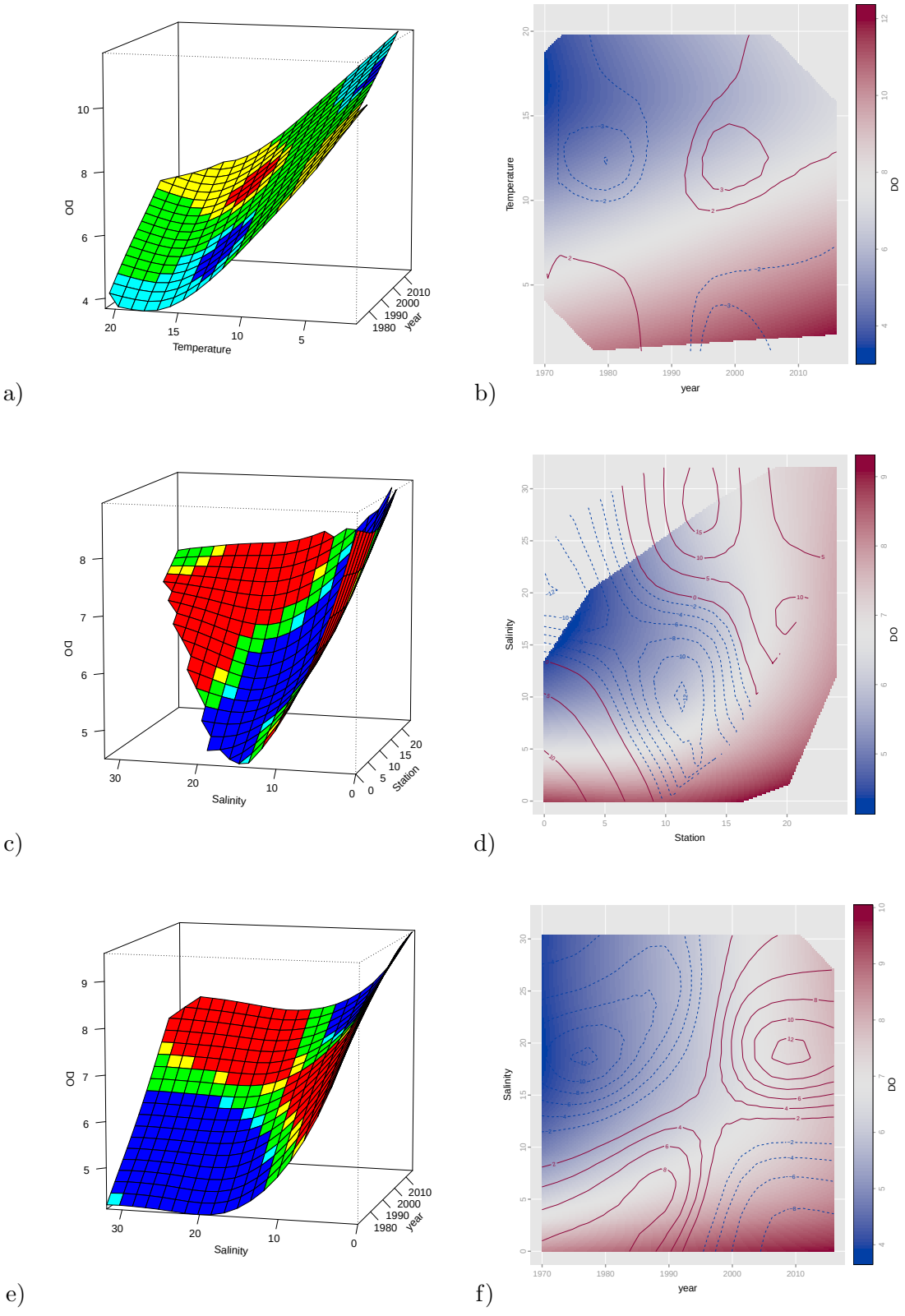
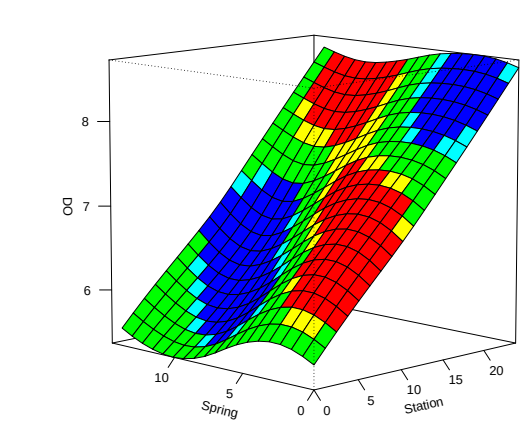


FIGURE 5.9: Interaction plots of a-b) Day of Year : Station, c-d) Day of Year : Year, and e-f) Temperature : Station for complex additive mixed model.

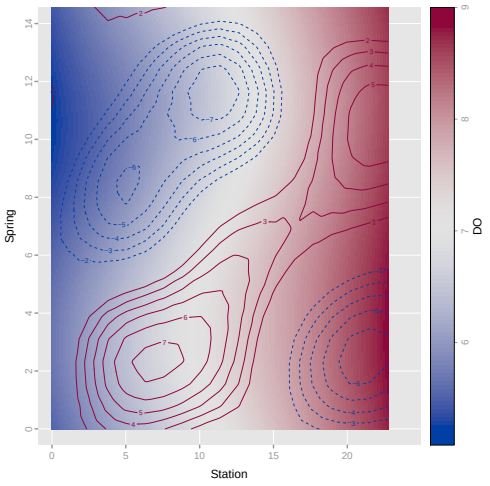


a) b) c) d) e) f)

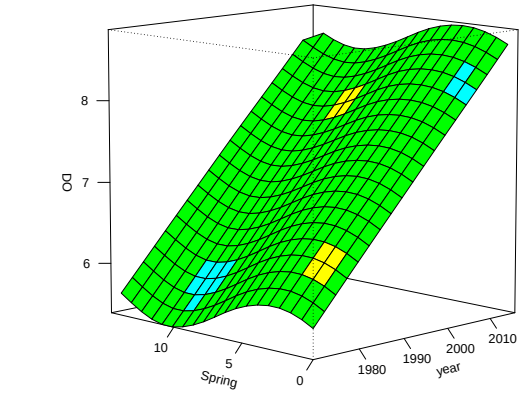
FIGURE 5.10: Interaction plots of a-b) Temperature : Year, c-d) Salinity : Station, and e-f) Salinity : Year for complex additive mixed model.



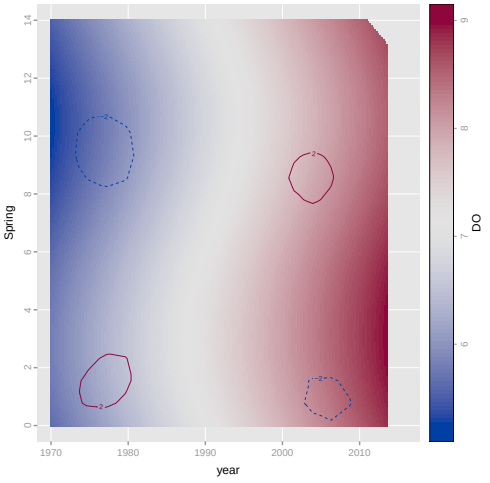
a)



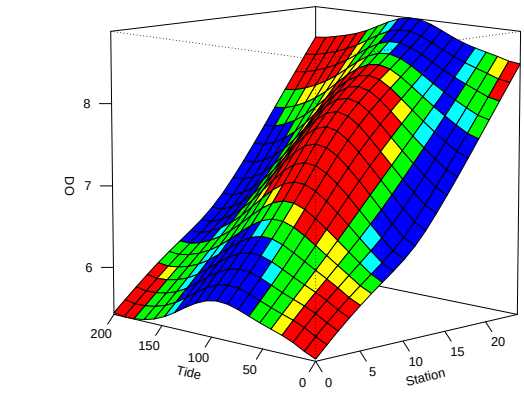
b)



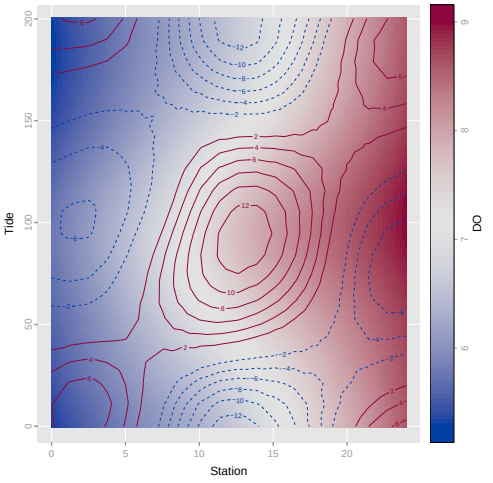
c)



d)



e)



f)

FIGURE 5.11: Interaction plots of a-b) Spring : Station, c-d) Spring : Year, and e-f) Tide : Station for complex additive mixed model.

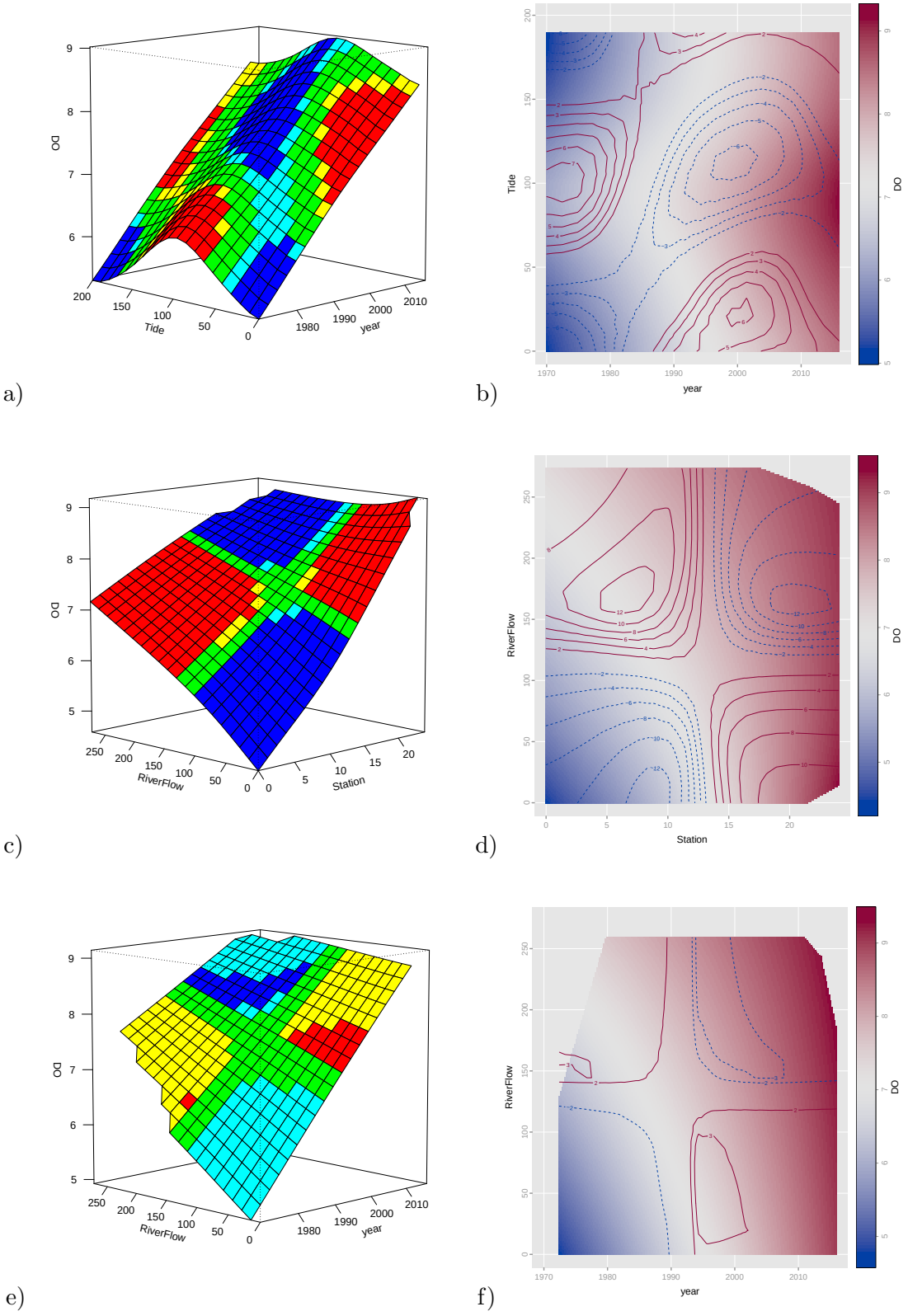
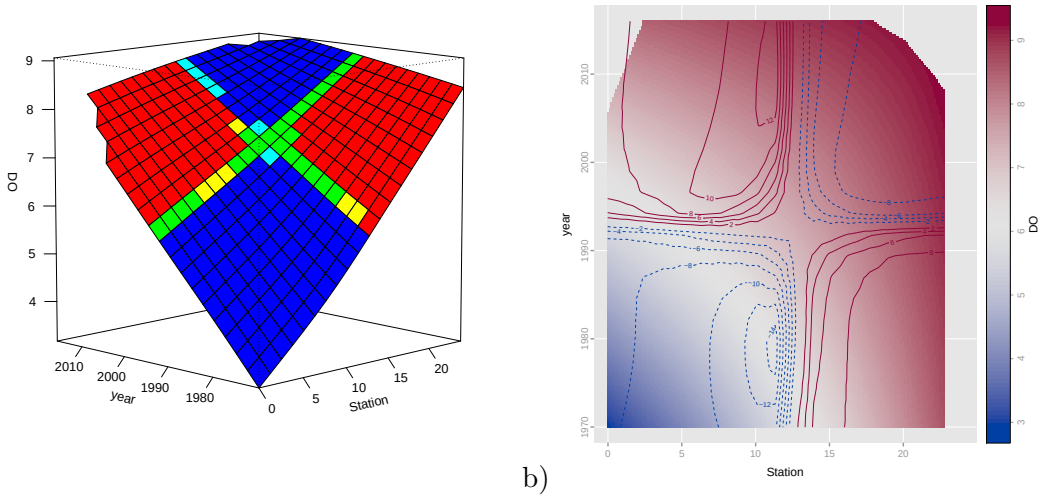


FIGURE 5.12: Interaction plots of a-b) Tide : Year, c-d) River Flow : Station e-f) River Flow : Year for complex additive mixed model.



a) b)

FIGURE 5.13: Interaction plots of a-b) Year : Station for complex additive mixed model.

5.1.4.3 Trivariate Interaction Terms

Where univariate main effects can be depicted as curves in a 2-dimensional space and bivariate interactions terms can be depicted as surfaces in a 3-dimensional space via perspective or contour plots, depiction of trivariate interaction terms can be abstractly interpreted as 3-dimensional objects in a 4-dimensional space. This can be done by using the `R` package `rpanel` which produces dynamic perspective and contour plots. The dependent variable and two independent variables are represented as a surface just as they would be in a bivariate interaction term. The value of the third independent variable is then varied and the position and curvature of the surface are adjusted accordingly.

Using the Salinity : Year : Station interaction as an example for illustrative purposes, these adjustments can be seen in Figure 5.14. Each row of panels has a specific value for Salinity depicted by the light blue bar along the Salinity axis found at the top of the panels. The first row has Salinity=0, the second row has Salinity=5, etc. The left column of panels depict the perspective plot of the interaction alone at different values of Salinity. The middle column of panels depict the perspective plots of the interaction with lower terms included at different values of Salinity. The right column of panels depict the contour plots of of the interaction with lower terms included at different values of Salinity. The contour plots are the most telling, as they combine the information given by the two corresponding perspective plots. The trivariate interaction terms can be interpreted in the same way as the bivariate interaction terms; the colouring behind the contours represent the vertical position of the surface in the middle panel and the contours themselves represent the position of the surface in the left panel. Each contour

within the contour plots depict the adjustment made by the interaction in terms of standard errors.

As mentioned above, Figure 5.14 depicts the Salinity : Station : Year interaction to help explain the interpretation of the trivariate interaction plots. This interaction, together with Salinity as the independent variable across the top axis, displayed the most dramatic and noticeable animation compared to the other 5 trivariate interactions. However, a much more interpretable depiction would have the Year independent variable across the top axis as is the case in Figures 5.15-5.20. Allowing the Year to be varied shows how the interactions change over time. Once again the lower terms are included in these contour plots so there is a sense of the overall dissolved oxygen levels depicted by the underlying colours and contour lines depicting the adjustments made by the interactions.

Figures 5.15-5.20 are composed of 6 panels each. Each panel represents a snapshot of the total dissolved oxygen and the interaction at a particular time. The values of Year represented are 1970, 1980, 1990, 2000, 2010, and 2016. All 6 trivrite interaction plots show a significant increase in the dissolved oxygen over the years, as the underlying colour tends from blue to red. This is most evident upstream at the smaller values of Station. There are also distinct and significant transformations of the interaction contours that are worth noting. It seems there exist specific coordinates where maximums transform into minimums and vice versa as Year is varied from 1970 to 2016. The approximate location of these coordinates at Year 1970 are as follows:

- (Day of Year, Station) — (120,0) (300,2) (130,15) (300,18) — Figure 5.15
- (Temperature, Station) — (2,12) (11,8) — Figure 5.16
- (Salinity, Station) — (10,0) (10,16) (30,14) — Figure 5.17
- (Spring, Station) — (3,12) (12,10) — Figure 5.18
- (Tide, Station) — (0,12) (20,0) (100,14) (130,0) — Figure 5.19
- (River Flow, Station) — (20,12) (250,12) — Figure 5.20

It is hard to believe this behaviour is purely coincidental. Some reasons for this behaviour that come to mind are the changing morphology due to sediment transport or the absence of dredging in the last few decades, although this is speculation.

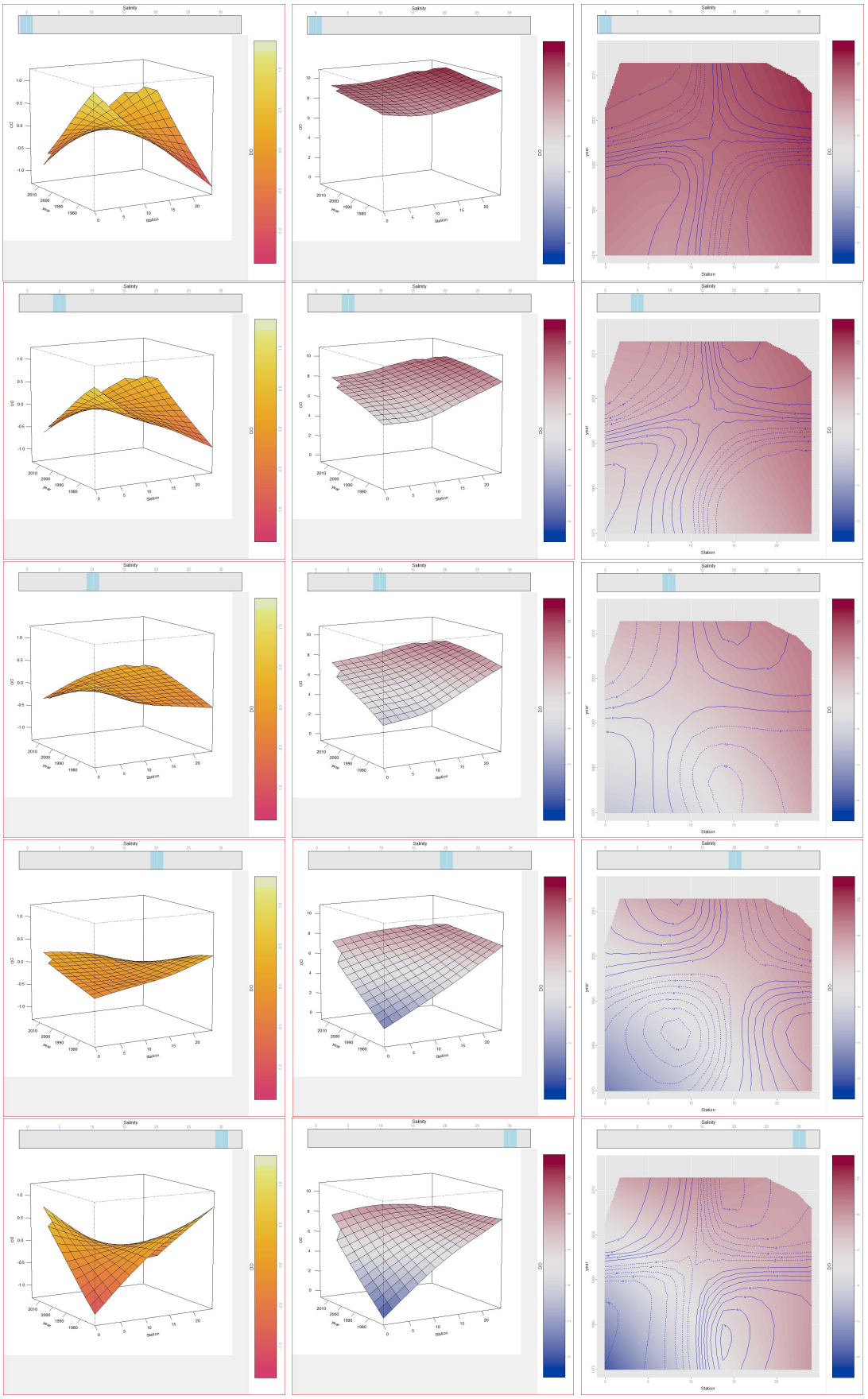


FIGURE 5.14: Plots of Salinity : Year: Station interaction term for complex additive mixed model. The left column depicts the interaction alone. The middle and right columns depict the interaction with lower terms included.

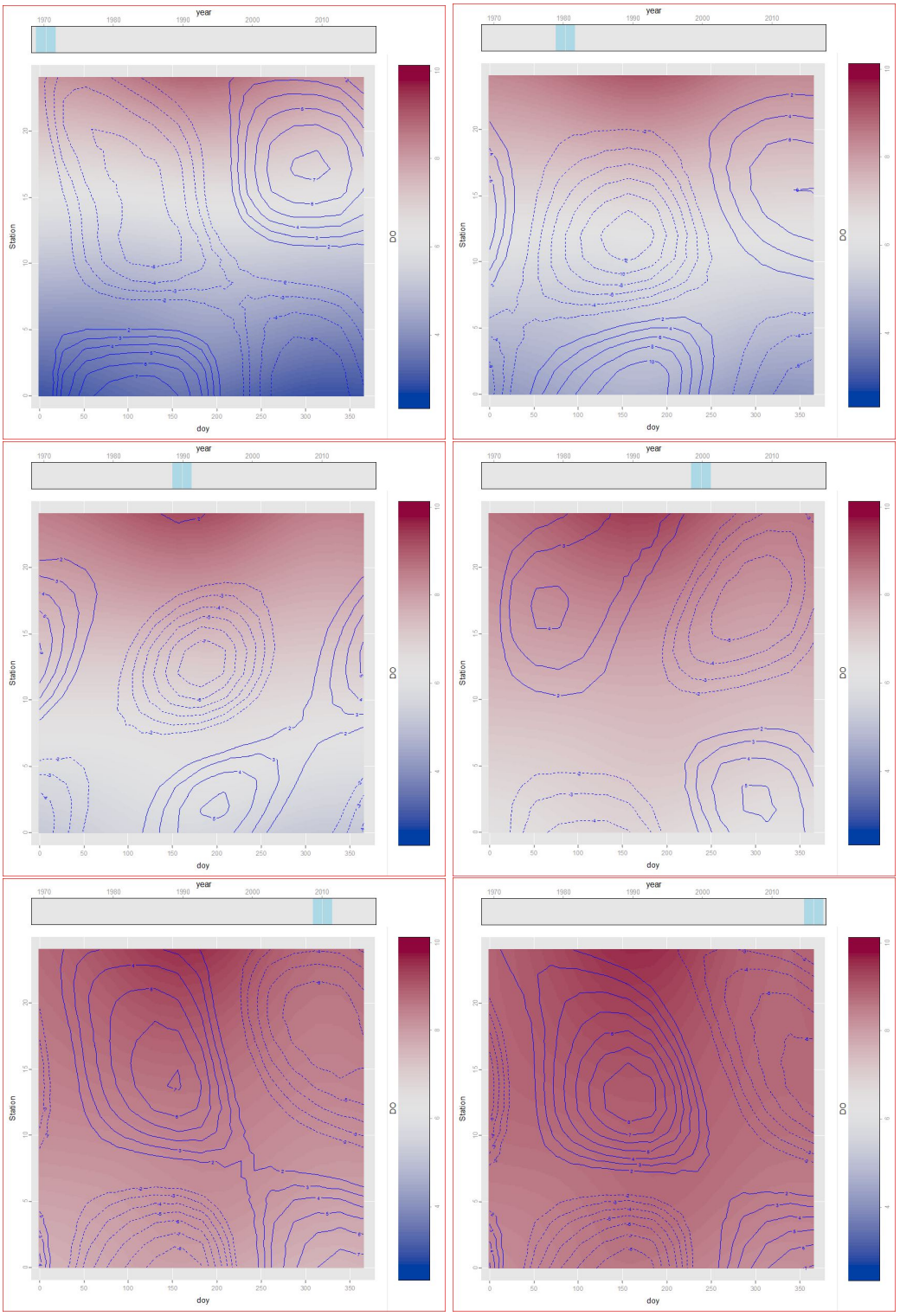


FIGURE 5.15: Plots of Day of Year : Station : Year interaction term of complex additive mixed model.

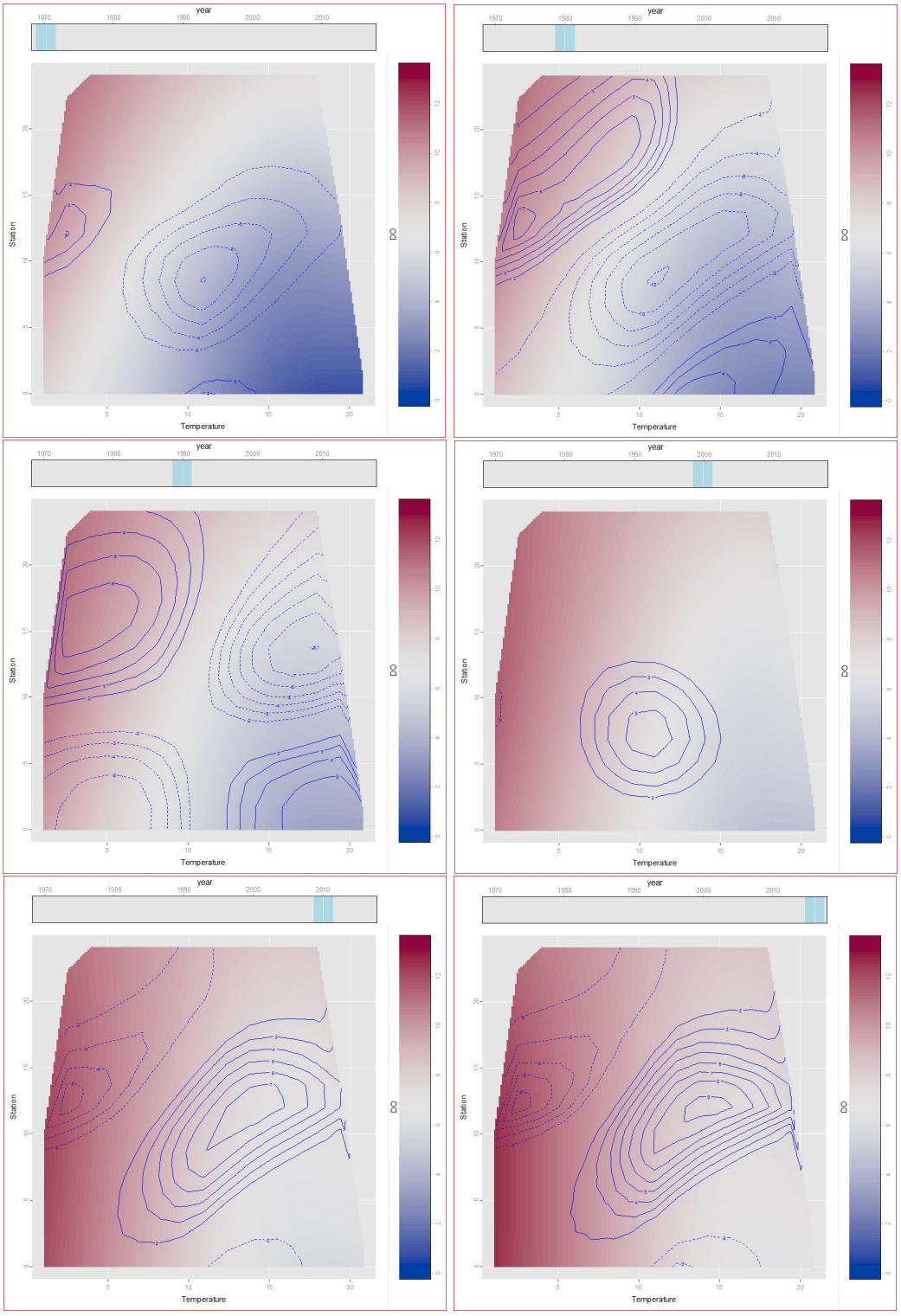


FIGURE 5.16: Plots of Temperature : Station : Year interaction term of complex additive mixed model.

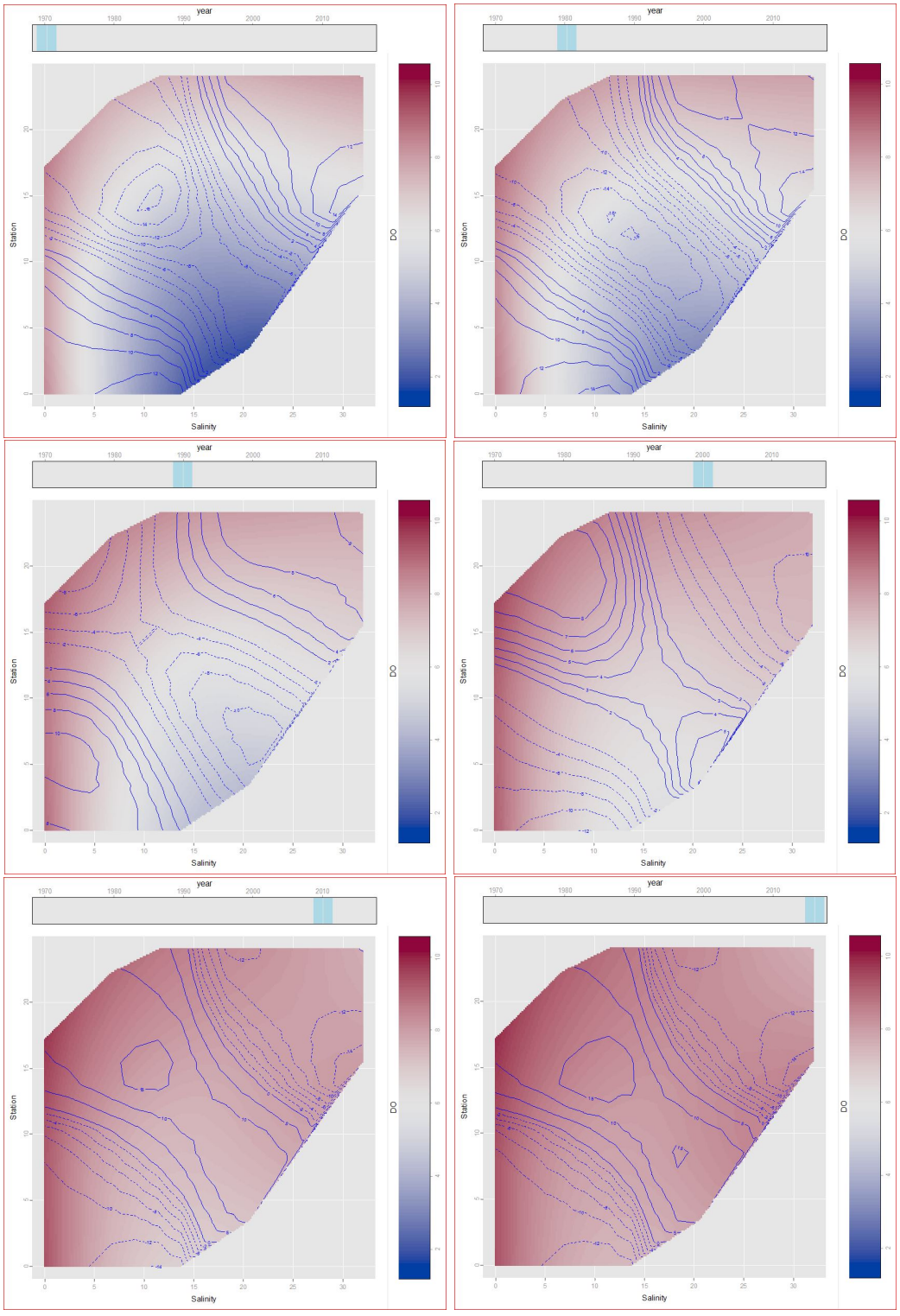


FIGURE 5.17: Plots of Salinity : Station : Year interaction term of complex additive mixed model.

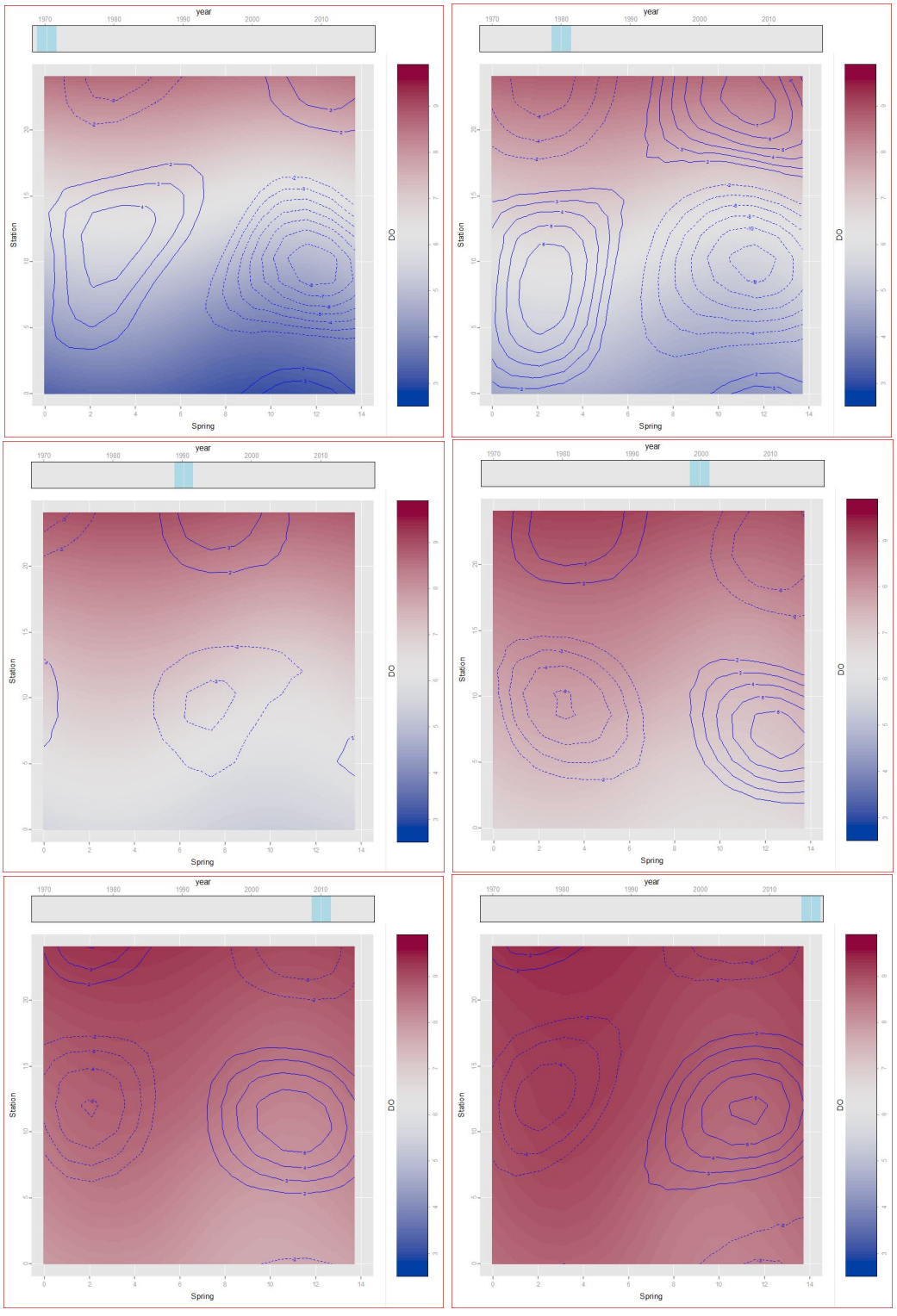


FIGURE 5.18: Plots of Spring : Station : Year interaction term of complex additive mixed model.

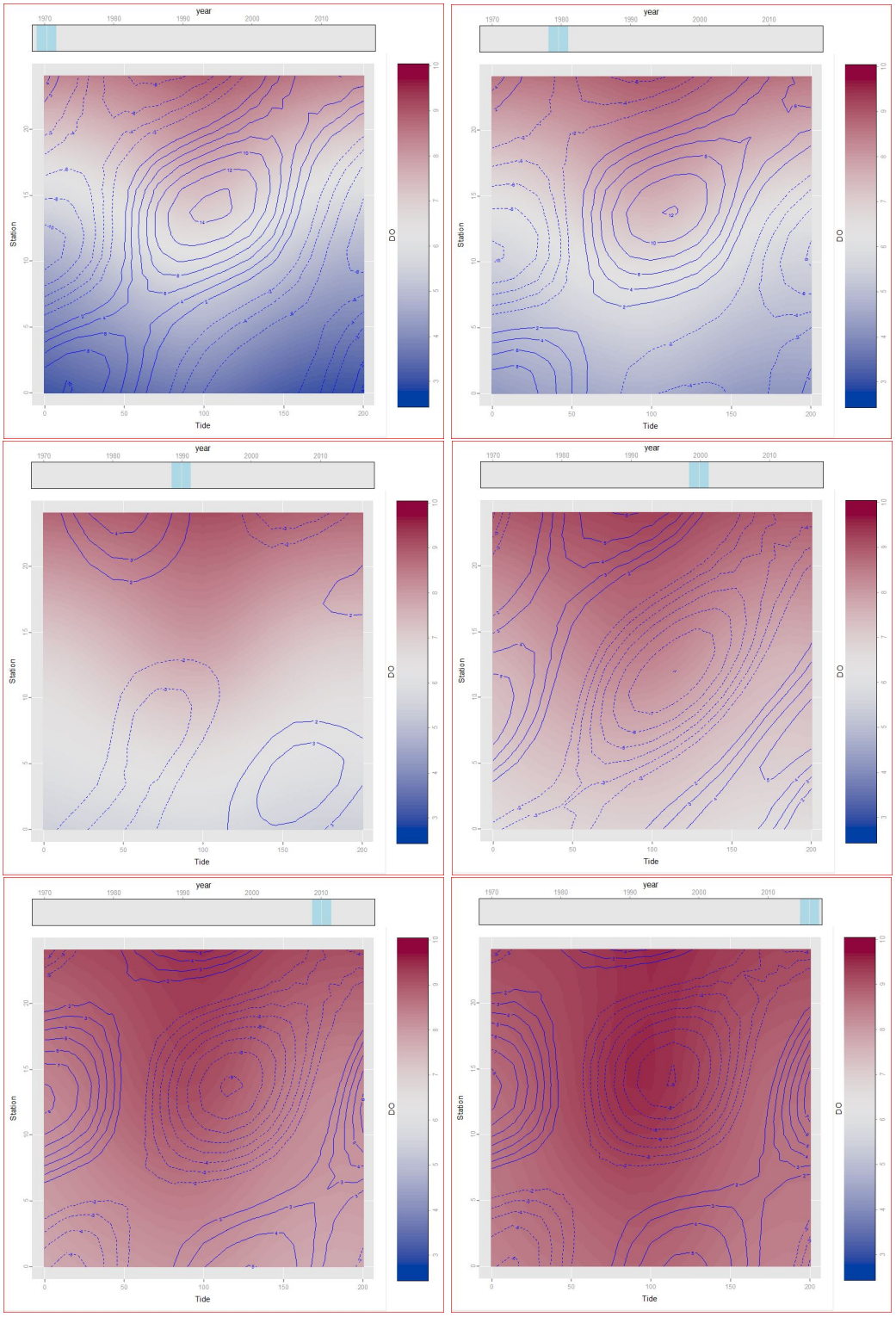


FIGURE 5.19: Plots of Tide : Station : Year interaction term of complex additive mixed model.

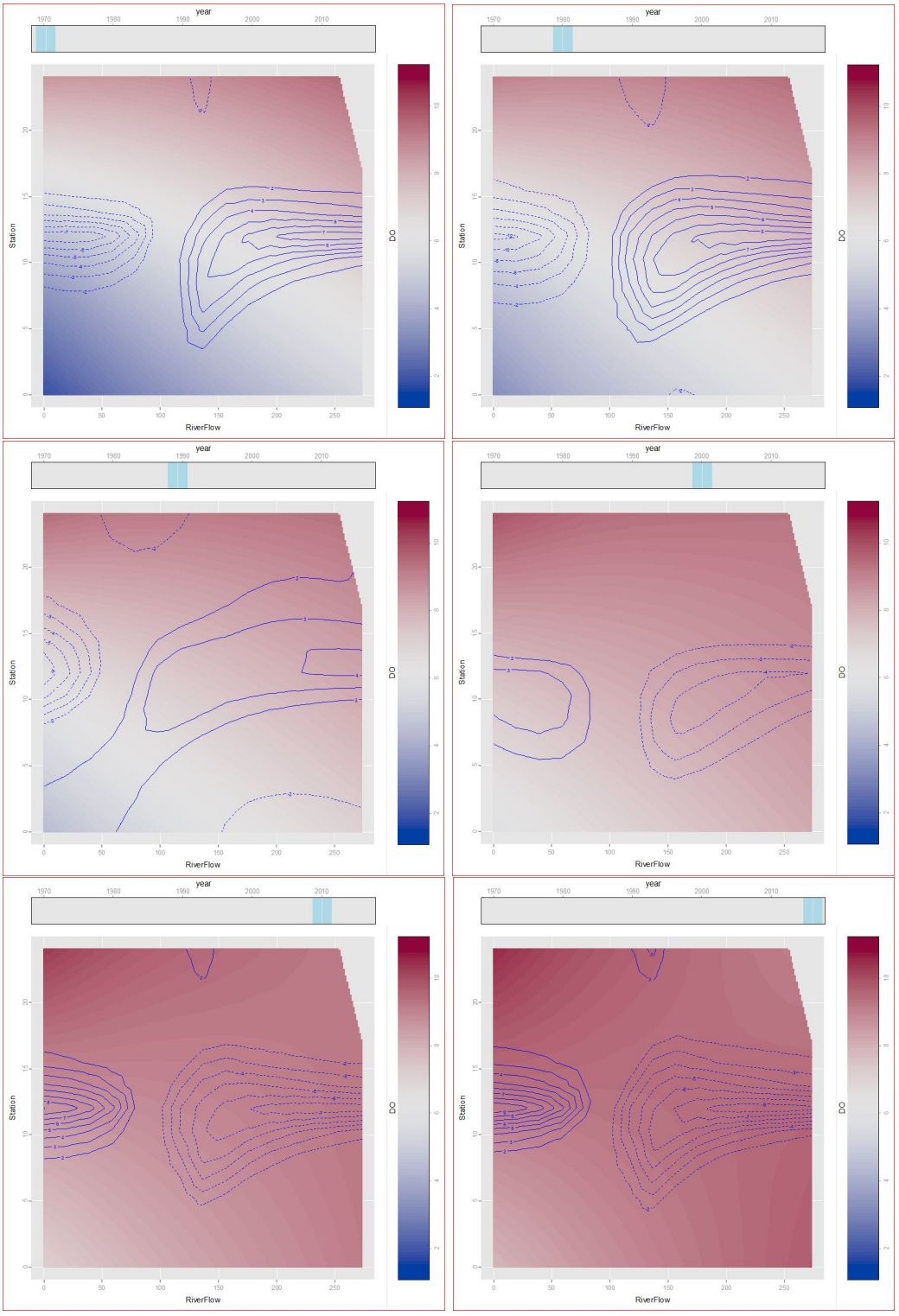
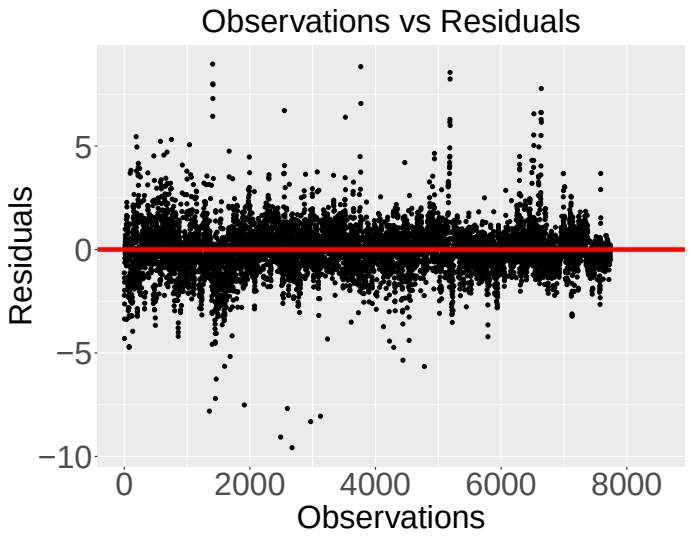


FIGURE 5.20: Plots of River Flow : Station : Year interaction term of complex additive mixed model.

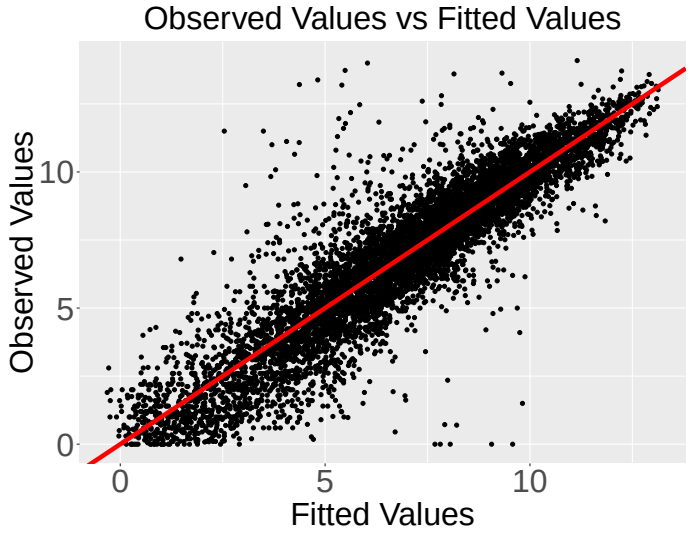
5.1.4.4 Final Model Performance

Although it is customary for model comparisons to be made with a view to constructing a parsimonious model, this is not an aim of this thesis. More specifically, one aim of this thesis is to generate an additive mixed model with interactions for use as a tool in detecting sudden yet subtle changes in the dissolved oxygen in a system which has shown a steady improvement over the last 4 plus decades. The inclusion of extra terms deemed insignificant would likely not affect the model's ability to detect such changes. As for the random effects component, there is a strong argument to be made that the conditions in place which are not part of the explanatory covariates during a particular day of sampling would drive the dissolved oxygen up or down.

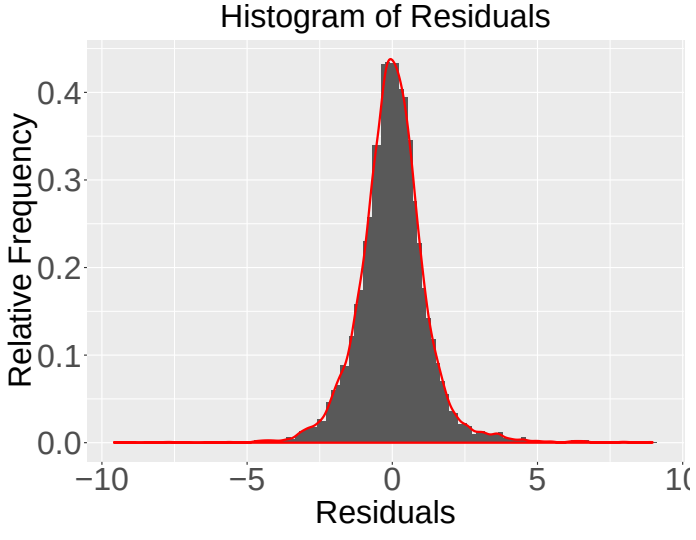
As the final additive mixed model with interactions is a flexible model, the usual linear assumptions are not in place. However, it is important to see if the residuals follow a random normal distribution. Panel a) of Figure 5.21 depicts the residuals of the raw data from their expected values in chronological order. These residuals show the desired random pattern centered about the horizontal red line representing a residual of 0. Panel b) depicts the observed values against the fitted values. The points again show the desired random pattern centered about the red line representing the location where fitted values equal observed values. There is no tendency for the points to fall systematically above or below the line. This shows the model does not exhibit bias at any particular predicted value. Panel c) depicts a histogram of the residuals. Here it is evident the residuals are normally distributed. Overall, the final additive mixed model with interactions fits the data well.



a)



b)



c)

FIGURE 5.21: Plots of residuals of final additive mixed model with interactions.

5.2 Discussion

Chapter 5 initially fitted a simple additive model with no interactions to the River Run data. The next step was to adjust the degrees of freedom for each of the main effects. Several smoothing parameter selection methods were applied. BIC managed to achieve minimum scores with reasonable corresponding degrees of freedom for all model covariates. GCV, AIC, and AICc were able to achieve minima in for some covariates, but the corresponding degrees of freedom were most often too large to seem plausible. BIC was then used to sequentially optimize the degrees of freedom for an additive model with no interactions and an additive mixed model with no interactions.

When attempting to fit an additive model with interactions, it was unknown if the order of the optimization of the main effects degrees of freedom had any influence on these final optimized degrees of freedom. A simulation study showed order of optimization was not of major influence to the final optimized degrees of freedom for simple bivariate additive models with interactions. However, the same might not be true for more complex additive mixed models with interactions. This was investigated further later in the chapter.

An additive mixed model with interactions was then fitted to the River Run data. This model had bivariate and trivariate interactions involving Year and Station. The main effects degrees of freedom were optimized in the same order that was used for the simple additive model with no interactions. Furthermore, 3 other permutations of degrees of freedom optimization were also tried. The result was that the degrees of freedom changed for different permutation, but the change was small. The original permutation yielded the smallest degrees of freedom. This permutation was used for the final additive mixed model with interactions. The main effects and interactions were depicted graphically and interpreted. It is not the aim of this thesis to remove any insignificant model terms, but it is reassuring that very few interactions showed adjustments small enough to be deemed insignificant. Finally, plots of the model residuals were presented and showed these residuals were random normal in nature.

The complex additive mixed model with interactions is representative of the River Run data. However, it is not very useful in detecting sudden subtle increases in the dissolved oxygen in a system which is already undergoing a gradual overall improvement. A better approach may be found in the increased sensitivity of partial derivatives.

Chapter 6

Derivative Estimation

The analysis presented in the previous chapter confirms that the River Clyde has shown a steady improvement in the dissolved oxygen levels in the last few decades. This is evident upon reviewing the River Run data set collected by SEPA since 1970. There is interest in seeing if particular upgrades to certain wastewater treatment facilities have caused subtle yet distinctive improvements in the water quality. This may be expressed in a faster rate of improvement at the times of interest. Estimation of the partial derivative of the dissolved oxygen with respect to time evaluated at specific times and at specific locations along the river provides the motivation for this chapter.

6.1 Literature Review

There are many methods for evaluating the impact of particular events in a stationary system. For example, there is a very extensive body of work on the detection of change points which could be attributed to the events of interest. Horváth and Rice (2014) provide a good point of entry to this literature by surveying frequently used methods, detailing developments of these methods, and providing extensions to some of the results. Killick et al. (2012) show how computational efficiency, which is linear in the number of observations, can be achieved with large sample sizes and Wang and Samworth (2018) discuss the case where the response is highly multivariate. The present situation is different for three reasons. Firstly, the impacts of specific events are being sought within the context of a complex regression model with multiple covariates and complex interactions. There is the potential of having a model with eight main effects, thirteen bivariate interactions, and six trivariate interactions. Secondly, the system being studied is already exhibiting general improvement. This steady increase in dissolved oxygen has been present over the 46 year span of the data. A third issue is that we

do not necessarily expect that the impact of the events of interest will be immediate. The waste water treatment upgrades, specifically at Dalmuir, were implemented over a 4 year span. Instead of a sudden 'switched on' effect, the impact is expected to be spread over a period of time although on a faster timescale than the general background improvement.

These considerations lead to the estimation of partial derivatives with respect to time as the mechanism for the detection of an increased rate of change, attributable to the treatment facility upgrades. Derivatives in models with a single covariate have a long history, as nonparametric estimates of a function can often be differentiated in a straightforward manner to provide an immediate estimate of the derivative. In an early paper, Müller et al. (1987) recognised that the derivative of a good estimate of a function does not necessarily provide a good estimate of the derivative of the function. To address this, these authors proposed a method for controlling smoothness based on a comparison of the estimate of the derivative with an empirical version constructed by differencing the data. Charnigo et al. (2011) took a similar but more general approach involving empirical derivatives but targeting a generalisation of the C_p criterion (Mallows, 1973). Aldrin (2006) considered the case of P-splines and advocated penalizing the slope and curvature in standard additive models, not in the context of derivative estimation, and found that penalizing both slope and curvature performed as good or better than penalizing curvature alone when estimating the underlying function.

Simpkin and Newell (2013) explored the use of P-splines with two smoothness penalties on differences in the context of derivative estimation and identified that improved performance could be achieved. In choosing the combination of penalty orders, the authors considered orders of (1,2), (1,3), and (2,3), and found little difference in performance, albeit order (1,2) performed slightly better over all. For this reason, the authors adopted penalty order (1,2) for the remainder of the study. In a more general setting, Heckman and Ramsay (2000) investigate the effects of replacing a smoothness penalty based on the usual functional of the second derivative, for which a linear function is the null model, with one constructed from more general linear differential operators to allow a wider range of reference models, as may suit particular applications.

Assessing the derivative and partial derivative estimation performance is the aim of this simulation study. The findings of Aldrin (2006) suggest that a slope and curvature penalty performs well for non-derivative estimation, while Simpkin and Newell (2013) advocate additive penalties specifically for estimating derivatives. This provides the motivation to analyze the performance of higher order additive penalties. It is reasonable to assume that a (2,3) penalty on the underlying function translates to a (1,2) penalty in the derivative. It also make sense to analyze the performance of a penalty of order 3

on the underlying function, which may behave as a penalty of order 2 on the derivative. A more detailed review of these two sources is given below.

6.1.1 Improved predictions penalizing both slope and curvature in additive models - Aldrin (2006)

We have seen many instances where the second derivative, or curvature, is exclusively penalized when constructing a smooth function. Other order derivatives can be used in place of the second derivative, as there is nothing special about the second derivative; Eilers and Marx (1996). Aldrin (2006) proposes adding an additional penalty on the first derivative, or slope, of the smooth function.

The slope and curvature penalized smooth function will aim to minimize an objective function of the form

$$S = \sum_{i=1}^{n}[y_i - f(x_i)]^2 + \lambda_1 \int [f^{(1)}(x)]^2 dx + \lambda_2 \int [f^{(2)}(x)]^2 dx \tag{6.1}$$

for the univariate case ($p = 1$) and

$$S = \sum_{i=1}^{n}[y_i - f(x_i)]^2 + \lambda_1 \sum_{j=1}^{p} \int [f_j^{(1)}(x)]^2 dx + \lambda_2 \sum_{j=1}^{p} \int [f_j^{(2)}(x)]^2 dx \tag{6.2}$$

for the p-variate case, where $\boldsymbol{x_i} = (x_{1_i}, ..., x_{p_i})^T$ and $f = f_1 + ... + f_p$. In the case of P-splines, the penalties on the first and second derivatives are replaced by first and second differences on the B-spline coefficients respectively. Thus, if we let k be the number of knots in each of the B-spline bases for all p expanitory covariates, objective function equation 6.2 becomes

$$S = \sum_{i=1}^{n}[y_i - f(\boldsymbol{x_i})]^2 + \lambda_1 \sum_{j=1}^{p}\sum_{l=2}^{k}(\alpha_{j_l} - \alpha_{j_{l-1}})^2 + \lambda_2 \sum_{j=1}^{p}\sum_{l=3}^{k}(\alpha_{j_l} - 2\alpha_{j_{l-1}} + \alpha_{j_{l-2}})^2 \tag{6.3}$$

or

$$S = \sum_{i=1}^{n}[y_i - f(\boldsymbol{x_i})]^2 + \lambda_1 \sum_{j=1}^{p}(\Delta^{(1)}\alpha_j)^2 + \lambda_2 \sum_{j=1}^{p}(\Delta^{(2)}\alpha_j)^2 \tag{6.4}$$

where $\Delta^{(1)}$ and $\Delta^{(2)}$ are defined in equations 3.12 and 3.13 respectively.

Aldrin (2006) compares the performance of three variations of the penalized slope and curvature method to five simpler methods of model construction. Those simpler methods are

- Mean - model with intercept only
- Ordinary least squares
- Ridge regression standardized by equal variance
- Penalty on just slope - λ_1 estimated while $\lambda_2 = \infty$
- Penalty on just curvature - λ_2 estimated while $\lambda_1 = 0$

The other three methods utilize double penalties, thus two smoothing parameters, λ_1 and λ_2, need to be selected. Aldrin (2006) mentions it is perhaps most natural to estimate both λ_1 and λ_2 for the double penalty model simultaneously, referring to it as DP12sim. The author also considers estimating the 2 tuning parameters sequentially. This is done by

- **Step 1** - Estimate λ_2 with fixed $\lambda_1 = 0$
- **Step 2** - Estimate λ_1 with λ_2 fixed to value found in **Step 1**.

This is referred to as DP12seq21. Aldrin (2006) also uses this method but starts with λ_1, calling it DP12seq12.

When it is time for smoothing parameter selection, the author uses a method he calls leave-k-out cross-validation. This method is similar to the bootstrap-smoothed cross-validation method referred to by Shao (1993) and Pan (1999). For a data set of n observations, this method defines k as

$$k \approx m/V \tag{6.5}$$

where m is the number of training data points drawn out without replacement and V is the number of groups of data defining a partition of the explanatory covariate of the training data. Assigning the first k observations to the first of the V groups, the second k observations to the second of the V groups, and so on, would be one way to create the partition. Selecting one of these V groups to leave out and constructing a model using the remaining $V - 1$ groups allows us to predict the k left out response variable values and compare them to the actual data points. This is repeated for all V groups. If $v = 1, ..., V$ denotes the different groups and the subscript $(-v)$ denotes the estimate without the v^{th} group, the criterion to be minimized is

$$\sum_{v=1}^{V} \sum_{i \in v} [y_i - \hat{f}_{(-v)}(\boldsymbol{x_i})]^2 \tag{6.6}$$

By permuting the n observations and repeating the partitioning process several times, we are able to take an average over expression 6.6. This method saves computation time compared to leave-one-out cross-validation. The leave-k-out cross-validation method also addresses problems with leave-one-out cross-validation which arise in the predictions using data outside the range of the training data. The comparison was made using RMSE defined by

$$\text{RMSE}^M = \sqrt{\frac{1}{S}\sum_{s=1}^{S}\frac{1}{n-m}\sum_{i\in D_s}[y_i - \hat{f}^M_{(-D_s)}(\boldsymbol{x_i})]^2} \tag{6.7}$$

where s is the simulation number, S is the number of simulations, M is the specific method , and $\hat{f}^M_{(-D_s)}(\boldsymbol{x_i})$ is the prediction using method M for the i^{th} observation based on the training data which excludes the observations in D_s. When comparing two methods, say method M under consideration and reference method R, Aldrin (2006) uses the log ratio

$$log\left(\frac{\text{RMSE}^M}{\text{RMSE}^R}\right) = log(\text{RMSE}^M) - log(\text{RMSE}^R). \tag{6.8}$$

When the ratio is negative, method M performs better than R and when the ratio is negative, method M performs worse than R.

After the simulation study, the author finds the leave-k-out cross-validation with permutation method was the optimal method when compared to leave-one-out and leave-k-out without permutation. Furthermore, the double penalty approach penalizing both slope and curvature (specifically DP12seq21) was as good as or generally outperformed the other methods considered. The instances where the slope and curvature penalty method under-performed were when the data sets were small.

6.1.2 An additive penalty P-Spline approach to derivative estimation- Simpkin and Newell (2013)

P-Splines are used to model relationships between explanatory and response variables which tend to be more complex than simpler linear relationships. The authors mention the use of P-Splines where derivative estimation is the primary goal in certain situations such as may arise in environmental and sports science. The authors then proceed to propose an additive penalty approach to derivative estimation by way of P-splines.

An attractive characteristic of B-splines is their derivatives can be calculated in terms of B-splines of lower order (De Boor et al. (1978)). Furthermore, Simpkin and Newell (2013) gives the equation for the first derivative of a q order B-spline with evenly spaced

knots

$$f^{(1)}(x) = h^{-1} \sum_{j=2}^{k} \Delta^{(1)} \alpha_j B_j(x; q-1) \tag{6.9}$$

where h is the distance between knots. Simpkin and Newell (2013) go on to give the general equation for the l-th derivative of a q order B-spline

$$f^{(d)}(x) = h^{-d} \sum_{j=d+1}^{k} \Delta^{(d)} \alpha_j B_j(x; q-d) \tag{6.10}$$

Other than Aldrin (2006), the authors mention Belitz and Lang (2008) as a past publication where a second penalty term is introduced for extra smoothing. In this publication Simpkin and Newell (2013) take an additive penalty approach for the specific task of derivative estimation. An additive penalty P - Spline with two penalty terms can be expressed as

$$\sum_{i=1}^{n} [y_i - f(x_i)]^2 + \lambda_1 \sum_{j=d_1+1}^{m} (\Delta^{(d_1)} \alpha_j)^2 + \lambda_2 \sum_{j=d_2+1}^{m} (\Delta^{(d_2)} \alpha_j)^2 \tag{6.11}$$

where n represents the number of observations, m represents the number of coefficients, and d_1 and d_2 represent the orders of the difference penalties. Once again using B as the B - Spline basis design matrix for the smooth function f and D as a differencing matrix, we aim to minimize

$$\|y - B\alpha\|^2 + \lambda_1 \left\|D_{d_1} \alpha\right\|^2 + \lambda_2 \left\|D_{d_2} \alpha\right\|^2$$

by taking the first derivative with respect to α and setting it equal to zero. The resulting estimate $\hat{\alpha}$ becomes

$$\hat{\alpha} = (B^T B + \lambda_1 D_{d_1}^T D_{d_1} + \lambda_2 D_{d_2}^T D_{d_2})^{-1} B^T y. \tag{6.12}$$

The derivatives are calculated in the same way as in the case of the single penalty by using equation 6.10.

When selecting the two smoothing parameters, Simpkin and Newell (2013) cite Belitz and Lang (2008), opting for GCV over CV because of similar results and the computational efficiency of GCV. The authors also found, via simulation, no conclusive difference between simultaneous and sequential smoothing parameter selection and thus used sequential because of computational efficiency.

Simpkin and Newell (2013) address the construction of variability bands for the derivative estimate when using both single penalty and additive penalty P-splines. The authors

first give a summary of variability bands for P-splines, citing Eilers and Marx (1996), for the underlying function f. The estimate of f can be expressed as

$$\hat{\boldsymbol{f}} = \boldsymbol{B}\hat{\boldsymbol{\alpha}} = \boldsymbol{H}\boldsymbol{y} \tag{6.13}$$

where the hat matrix $\boldsymbol{H}$ for the single penalty is calculated using the equation

$$\boldsymbol{H} = \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B} + \lambda \boldsymbol{D_d}^T\boldsymbol{D_d})^{-1}\boldsymbol{B}^T \tag{6.14}$$

for some difference penalty d. The variance of $\hat{f}$ can then be shown to be

$$\text{var}(\hat{f}) = \sigma^2 \boldsymbol{H}\boldsymbol{H}^{\boldsymbol{T}} \tag{6.15}$$

The effective dimension of the model can be given by the trace of $\boldsymbol{H}$, and thus the variance of $\boldsymbol{y}$ can be estimated by

$$\hat{\sigma}^2 = \frac{\left\|\boldsymbol{y} - \hat{\boldsymbol{f}}\right\|^2}{n - tr(\boldsymbol{H})}. \tag{6.16}$$

The variability bands for the underlying function estimate with standard error se are then defined by

$$\hat{f} \pm z_c se[\hat{f}] \tag{6.17}$$

and can be approximated by

$$\hat{f} \pm z_c \hat{\sigma} \sqrt{\text{diag}(\boldsymbol{H}\boldsymbol{H}^{\boldsymbol{T}})} \tag{6.18}$$

where z_c is the Z-score corresponding to the middle c percent .

This method of constructing variability bands is now extended to the function's l^{th} derivative estimates. The variability bands are given by

$$\hat{f}^{(l)} \pm z_c se[\hat{f}^{(l)}]. \tag{6.19}$$

Since we are interested in the first derivative, the standard error of first derivative needs to be calculated. Adapting to simpler notation for equation 6.9, the estimate for the first derivative can be expressed as

$$\hat{\boldsymbol{f}}' = h^{-1}\boldsymbol{B}'\boldsymbol{D_1}\hat{\boldsymbol{\alpha}}. \tag{6.20}$$

where $\boldsymbol{B'}$ is the B-spline derivative matrix and $\boldsymbol{D_1}$ is the difference matrix described in equation 3.14. The variance of $\hat{\boldsymbol{f'}}$ can be shown to be

$$var(\hat{\boldsymbol{f'}}) = h^{-2}\boldsymbol{B'D_1CC}^T\boldsymbol{D_1}^T(\boldsymbol{B'})^T\sigma^2 \tag{6.21}$$

where

$$\boldsymbol{C} = (\boldsymbol{B}^T\boldsymbol{B} + \lambda_1\boldsymbol{D}^T\boldsymbol{D})^{-1}\boldsymbol{B}^T \tag{6.22}$$

The variability bands can be calculated using equation 6.19 with $l = 1$ with se of $\hat{f}'$ expressed as

$$se[\hat{\boldsymbol{f'}}] = h^{-1}\hat{\sigma}\sqrt{\text{diag}(\boldsymbol{B'D_1CC}^T\boldsymbol{D_1}^T(\boldsymbol{B'})^T)} \tag{6.23}$$

This method of constructing variability bands for the first derivative estimate by way of single penalty P-splines can now be extended to additive penalty P-splines by replacing $\boldsymbol{H}$ in equation 6.14 with

$$\boldsymbol{H} = \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B} + \lambda_1\boldsymbol{D_{d_1}}^T\boldsymbol{D_{d_1}} + \lambda_2\boldsymbol{D_{d_2}}^T\boldsymbol{D_{d_2}})^{-1}\boldsymbol{B}^T \tag{6.24}$$

where the $\boldsymbol{D}$'s are the difference matrices of orders corresponding to the respective subscripts.

The authors used one level of added noise ($\sigma = \frac{1}{3} \times \text{range}(f)$) and a fixed number (50) of observations for the simulated data for the derivative performance portion of the study. The results show the additive penalty model outperformed the single penalty model in estimating the first derivatives of all 6 functions used in their simulation study. Here, more functions across a wider range of penalty order combinations will be used. Multiple levels of random noise and observations of 500 and 5000 will also be incorporated, as the number of observations for the River Run is in excess of 5000.

6.2 Justification for Penalty Order 3 and 2/3

The customary smoothing penalty imposed on the integrated square of a smooth function's second derivative suffices for the estimate of the underlying function. This method translates to second differences on the B-spline coefficients in the context of P-splines. However, when derivative estimation of a smooth function f is the primary goal, a penalty on integrated square of the third derivative of f would translate to a second order penalty of f'. Likewise, when implementing the additive penalties of Aldrin (2006), penalty orders of 2 and 3 for f would translate to penalty orders of 1 and 2 for f'.

Let $\boldsymbol{f} = \boldsymbol{B\alpha}$ for some B-spline design matrix $\boldsymbol{B}$ of degree q and vector $\boldsymbol{\alpha}$ of length m of corresponding coefficients. Now let $\boldsymbol{g} = \boldsymbol{f'}$. Equation 6.9 allows us to express $\boldsymbol{g}$ as

$$\boldsymbol{g} = h^{-1}\boldsymbol{B'\gamma} \tag{6.25}$$

where $\boldsymbol{\gamma} = \Delta\boldsymbol{\alpha}$ is a $m-1$ vector of coefficients and $\boldsymbol{B'}$ is the B-spline design matrix of degree $q-1$. We are interested in estimating $\boldsymbol{g}$. A curvature penalty, P, imposed on $\boldsymbol{g}$ would have the form

$$P = \lambda h^{-3} \int (g''(x))^2 dx. \tag{6.26}$$

The P-spline equivalent objective function would have the form

$$\begin{aligned} P &= \lambda h^{-3} \sum_{j=2}^{m} (\Delta^{(2)}\gamma_j)^2 \\ &= \lambda h^{-3} \sum_{j=2}^{m} (\Delta^{(2)}(\Delta\alpha_j))^2 \\ &= \lambda h^{-3} \sum_{j=3}^{m} (\Delta^{(3)}\alpha_j)^2 \end{aligned} \tag{6.27}$$

This shows a second order penalty on the derivative estimate corresponds to a third order penalty on the original function. A similar argument can be made for the combination of a 2 and 3 penalty order on the underlying function translating to a combination of a 1 and 2 order penalty on the derivative. Furthermore, when $q = 3$, the integrated squared error of the third derivative of $\boldsymbol{f}$ is

$$\begin{aligned} \int [\boldsymbol{f'''}(\boldsymbol{x})]^2 dx &= \int \sum_{j=3}^{m}\sum_{k=3}^{m} \Delta^{(3)}\alpha_j B_j(x;0)\Delta^{(3)}\alpha_k B_k(x;0)dx \\ &= \sum_{j=3}^{m}\sum_{k=3}^{m} \Delta^{(3)}\alpha_j\Delta^{(3)}\alpha_k \int B_j(x;0)B_k(x;0)dx \\ &= \sum_{j=3}^{m} (\Delta^{(3)}\alpha_j)^2 \int B_j^2(x;0)dx. \end{aligned} \tag{6.28}$$

The last line of equation 6.28 is true because B-splines of degree 0 do not overlap.

6.3 Simulation Study

Comparing the performance of various single penalty orders and combinations of double additive penalty orders in derivative and partial derivative estimation is the main aim of this simulation study. The additive penalty approach suggested by Simpkin and Newell

(2013) involves separate smoothing parameters for each penalty order. Because of the complexity of fitting a model with so many main effects and interactions, here a single smoothing parameter over both penalty orders will be applied. Allowing for separate smoothing parameters will, because of flexibility, necessarily produce better derivative estimates, but that approach can be the subject of future research.

There is also interest in seeing the effect of degrees of freedom, the level of random noise added, and the number of points in the simulated data set on the accuracy of the derivative estimates.

6.3.1 Simulation Design

The first derivatives and partial derivatives are of primary concern here. The first derivatives are estimated using equation 6.9 and compared to the actual first derivative. This comparison is made by the root mean squared error,

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{f}'(x_i) - f'(x_i))^2}{n}} \tag{6.29}$$

Known functions are fitted with P-splines using the various penalty orders on the variable which the derivative will be taken with respect to. These single and additive penalty orders are

- 1
- 2
- 1 and 2
- 1 and 3
- 2 and 3
- 3

The degrees of freedom are also varied from 3 to 9 within each penalty order group to see what effect this may have on the performance of the derivative estimate. The BIC score is calculated for each degree of freedom and a subsequent model is fitted with the optimal degrees of freedom for that particular simulation. To clarify, each simulation goes through the following steps:

1. A known function is generated using a random uniform independent covariate and random normal noise is added to the dependent covariate.

2. The first penalty order is selected.
3. The first degrees of freedom is selected.
4. A P-spline model is fitted
5. The RMSE and the BIC score are calculated.
6. The next degrees of freedom is selected
7. Repeat steps 4 through 6 until all values of degrees of freedom have been used.
8. A P-spline model is fitted using the value of degrees of freedom corresponding to the lowest BIC score.
9. The next penalty order is selected.
10. Repeat steps 3 through 9 until all penalty orders have been used.
11. End of one simulation. Repeat steps 1 through 10 for the next simulation.

The degrees of freedom effect may become a point of interest in future research regarding derivative estimation via P-splines.

It is reasonable to consider a wide variety of functions with different behaviors when trying to estimate derivatives and partial derivatives. Some of the functions used here are taken from Simpkin and Newell (2013), along with other univariate and bivariate functions. The univariate functions used by Simpkin and Newell (2013) are:

- $\boldsymbol{F_1}$ $y = x + 2e^{-16x^2}$ for $x \in [0, 1]$
- $\boldsymbol{F_2}$ $y = (sin(2\pi x^3))^3$ for $x \in [0, 1]$
- $\boldsymbol{F_3}$ $y = \sqrt{x(1-x)}sin(\frac{2\pi(1.25)}{(x+0.25)})$ for $x \in [0, 1]$
- $\boldsymbol{F_4}$ $y = \frac{1}{1+e^{-20(x-0.5)}}$ for $x \in [0, 1]$
- $\boldsymbol{F_5}$ $y = -cos(x - \frac{\pi}{2}) + 2e^{-16x^2}$ for $x \in [0, 1]$

The additional univariate functions are:

- $\boldsymbol{F_6}$ $y = sin(x)$ for $x \in [0, 2\pi]$
- $\boldsymbol{F_7}$ $y = sin(x)e^x$ for $x \in [0, 2\pi]$
- $\boldsymbol{F_8}$ $y = sin(x)e^{-x}$ for $x \in [0, 2\pi]$

- $\boldsymbol{F_9}$ $y = sin(x^2)$ for $x \in [0, 2\pi]$
- $\boldsymbol{F_{10}}$ $y = 1000(x - 0.2)(x - 0.3)(x - 0.5)(x - 0.7)(x - 0.9) + 1$ for $x \in [0, 1]$
- $\boldsymbol{F_{11}}$ $y = 100(x - 0.1)(x - 0.2)(x - 0.4)(x - 0.7)(x - 0.9) + 2$ for $x \in [0, 1]$

These univariate functions and their derivatives are depicted in Figures 6.1 - 6.11. The functions cover a wide variety of behaviors from very tame (e.g. $\boldsymbol{F_1}$, $\boldsymbol{F_4}$, and $\boldsymbol{F_7}$) to very erratic (e.g. $\boldsymbol{F_2}$, $\boldsymbol{F_3}$, and $\boldsymbol{F_9}$).

Bivariate functions were constructed for the purpose of evaluating the performance of the estimation of partial derivatives with respect to each of the explanatory variables. The additional bivariate functions are:

- $\boldsymbol{F_{12}}$ $y = sin(x_1)\sqrt{x_2}$ for $x_1 \in [0, 2\pi]$ and $x_2 \in [1, 2]$
- $\boldsymbol{F_{13}}$ $y = sin(x_1)(x_2)^2$ for $x_1 \in [0, 2\pi]$ and $x_2 \in [1, 2]$

It seems reasonable to include bivariate functions composed of products of some combinations of functions $\boldsymbol{F_1}$ through $\boldsymbol{F_{11}}$. These functions are

- $\boldsymbol{F_{14}}$ $=$ $\boldsymbol{F_1}(x_1) \times \boldsymbol{F_4}(x_2)$
- $\boldsymbol{F_{15}}$ $=$ $\boldsymbol{F_{10}}(x_1) \times \boldsymbol{F_8}(x_2)$
- $\boldsymbol{F_{16}}$ $=$ $\boldsymbol{F_5}(x_1) \times \boldsymbol{F_7}(x_2)$
- $\boldsymbol{F_{17}}$ $=$ $\boldsymbol{F_6}(x_1) \times \boldsymbol{F_{11}}(x_2)$

These bivariate functions and their partial derivatives with respect to each variable are depicted with perspective and contour surface plots in Figures 6.12 - 6.17.

Each function was simulated with noise levels of 0.02, 0.05, 0.1, 0.2, 0.4, and 0.8. Sample sizes of both 500 and 5000 observation were used. The number of iterations for each simulation run was 500.

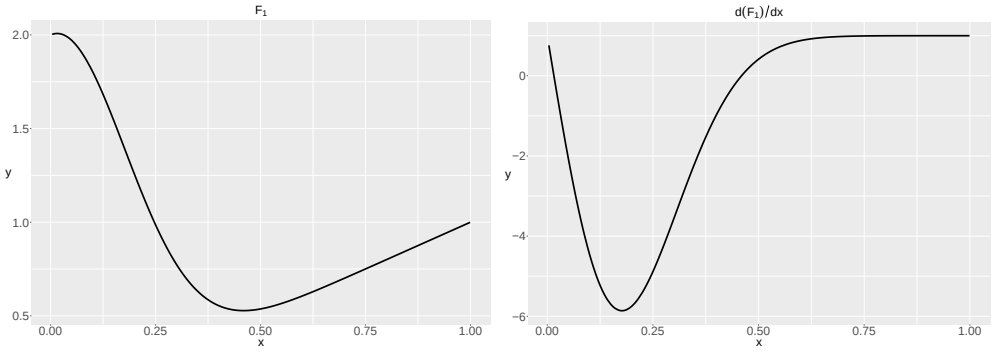


FIGURE 6.1: Plots of F_1 and $\frac{d(F_1)}{dx}$

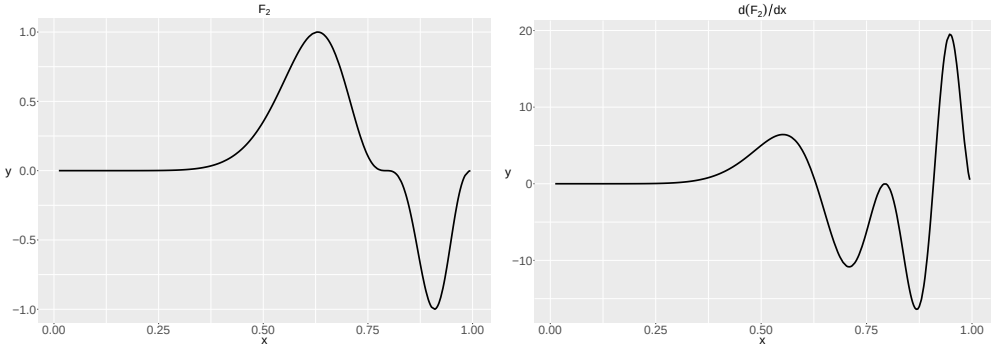


FIGURE 6.2: Plots of F_2 and $\frac{d(F_2)}{dx}$

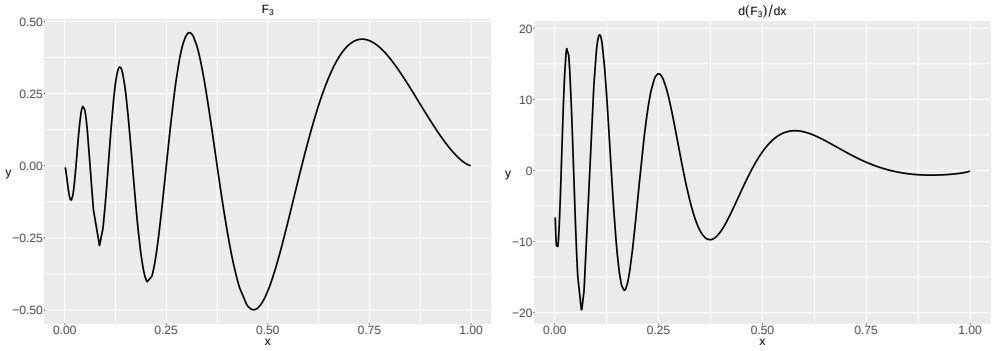


FIGURE 6.3: Plots of F_3 and $\frac{d(F_3)}{dx}$

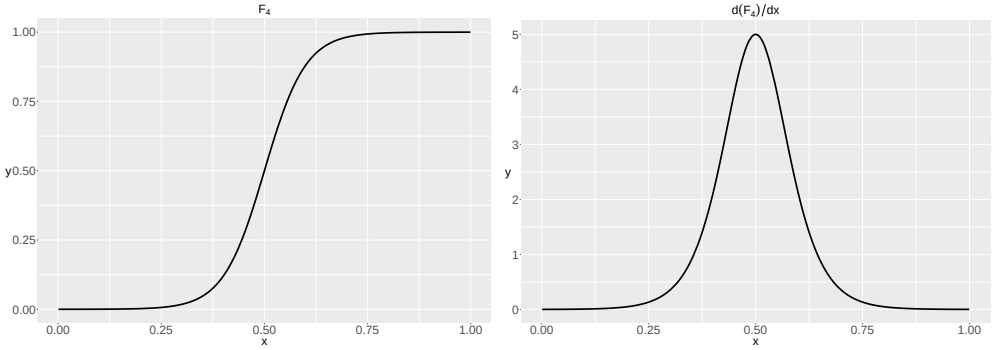


FIGURE 6.4: Plots of F_1 and $\frac{d(F_1)}{dx}$

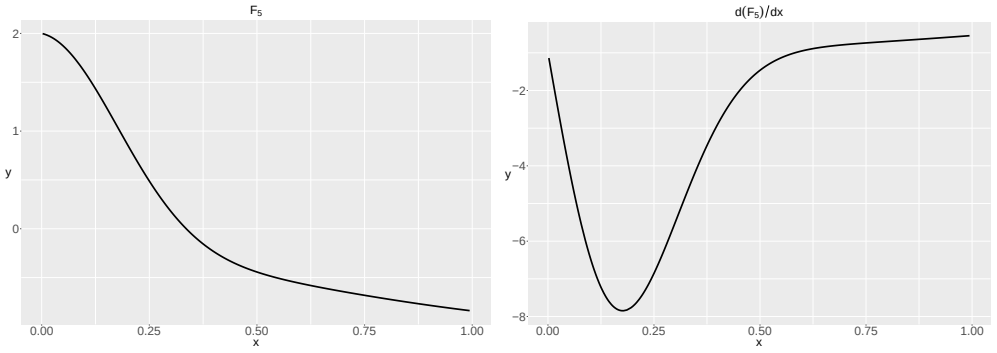


FIGURE 6.5: Plots of F_5 and $\frac{d(F_5)}{dx}$

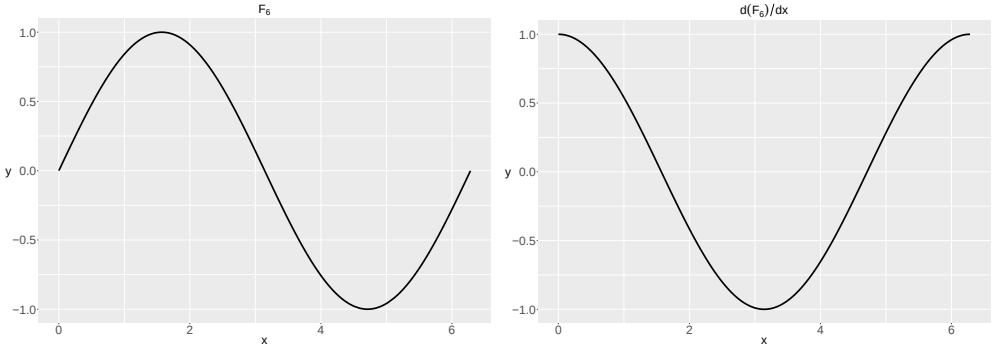


FIGURE 6.6: Plots of F_6 and $\frac{d(F_6)}{dx}$

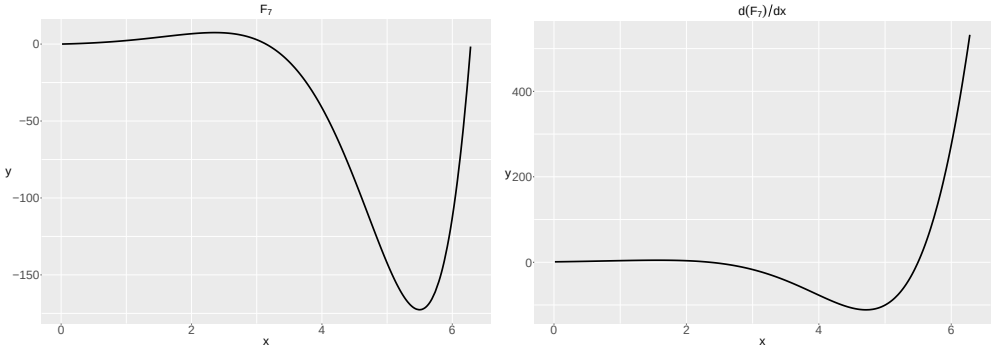


FIGURE 6.7: Plots of F_7 and $\frac{d(F_7)}{dx}$

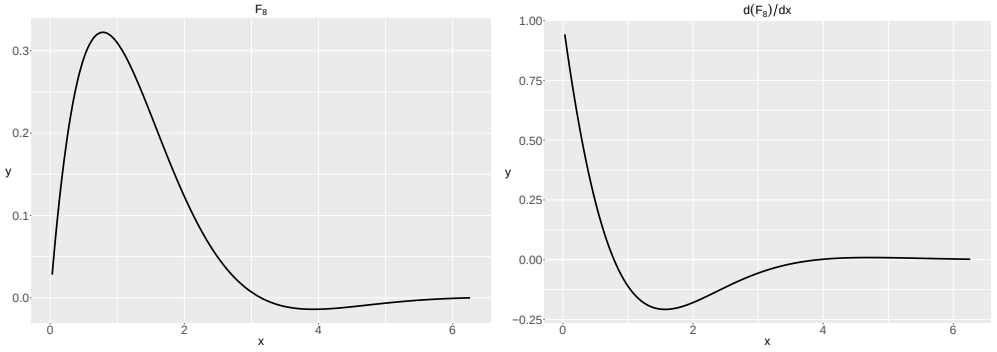


FIGURE 6.8: Plots of F_8 and $\frac{d(F_8)}{dx}$

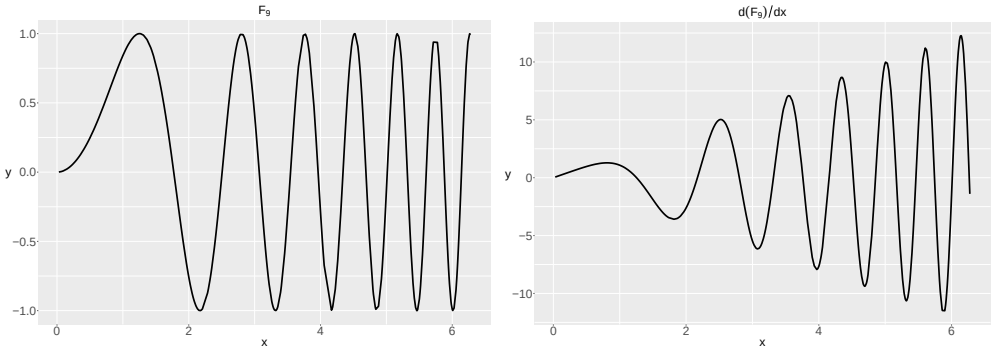


FIGURE 6.9: Plots of F_9 and $\frac{d(F_9)}{dx}$

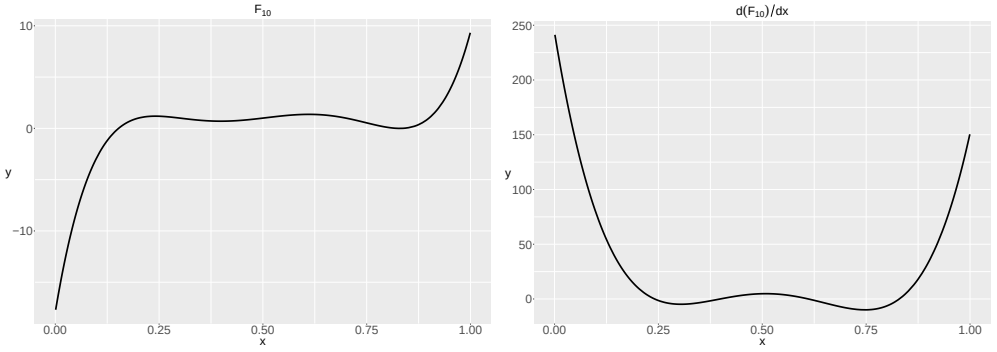


FIGURE 6.10: Plots of F_{10} and $\frac{d(F_{10})}{dx}$

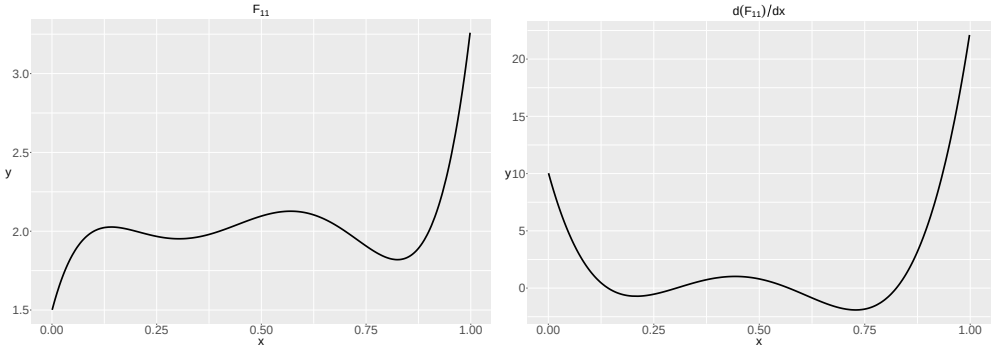


FIGURE 6.11: Plots of F_{11} and $\frac{d(F_{11})}{dx}$

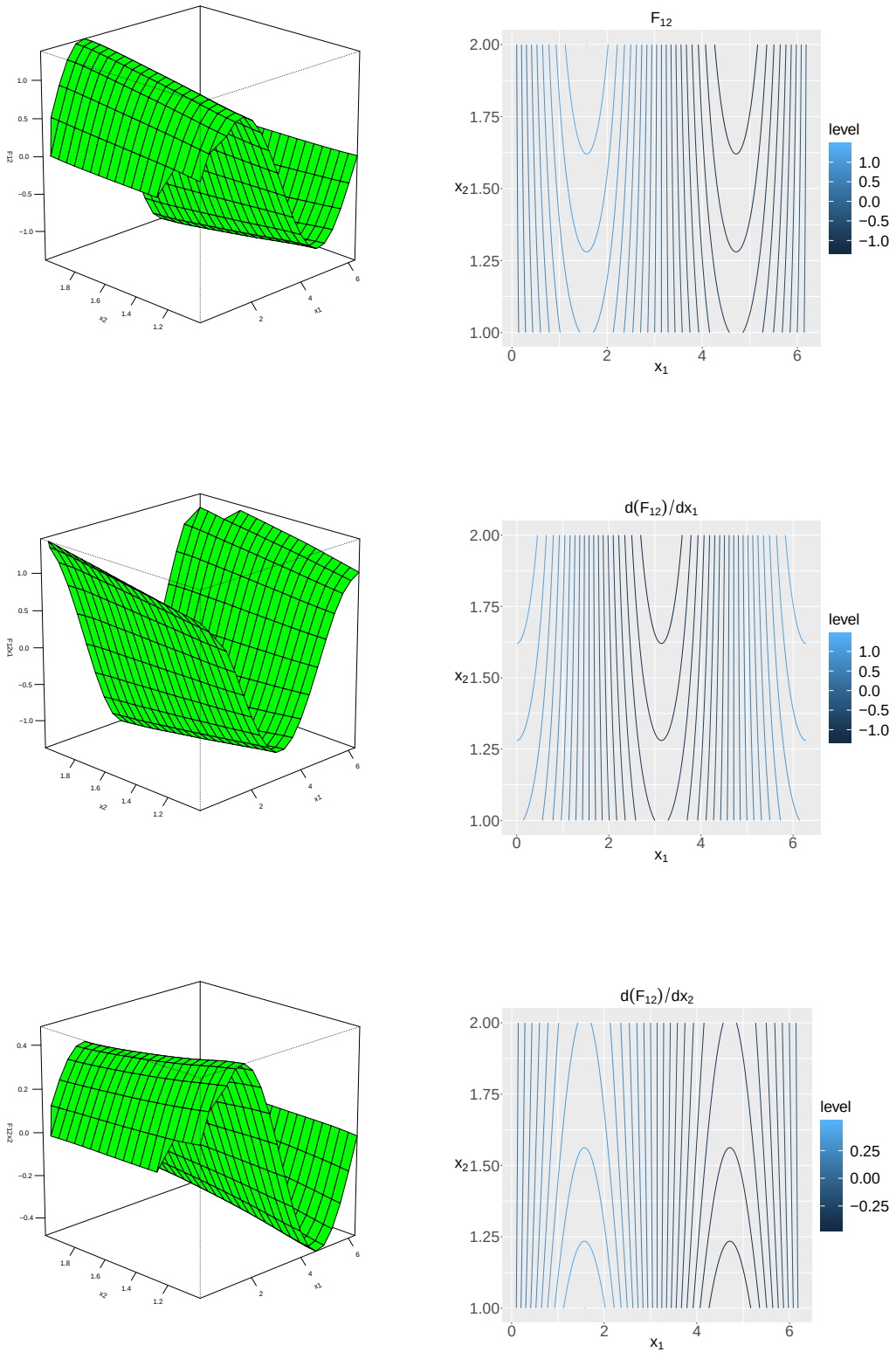


FIGURE 6.12: Perspective and contour plots of F_{12}, $\frac{\partial F_{12}}{\partial x_1}$, and $\frac{\partial F_{12}}{\partial x_2}$.

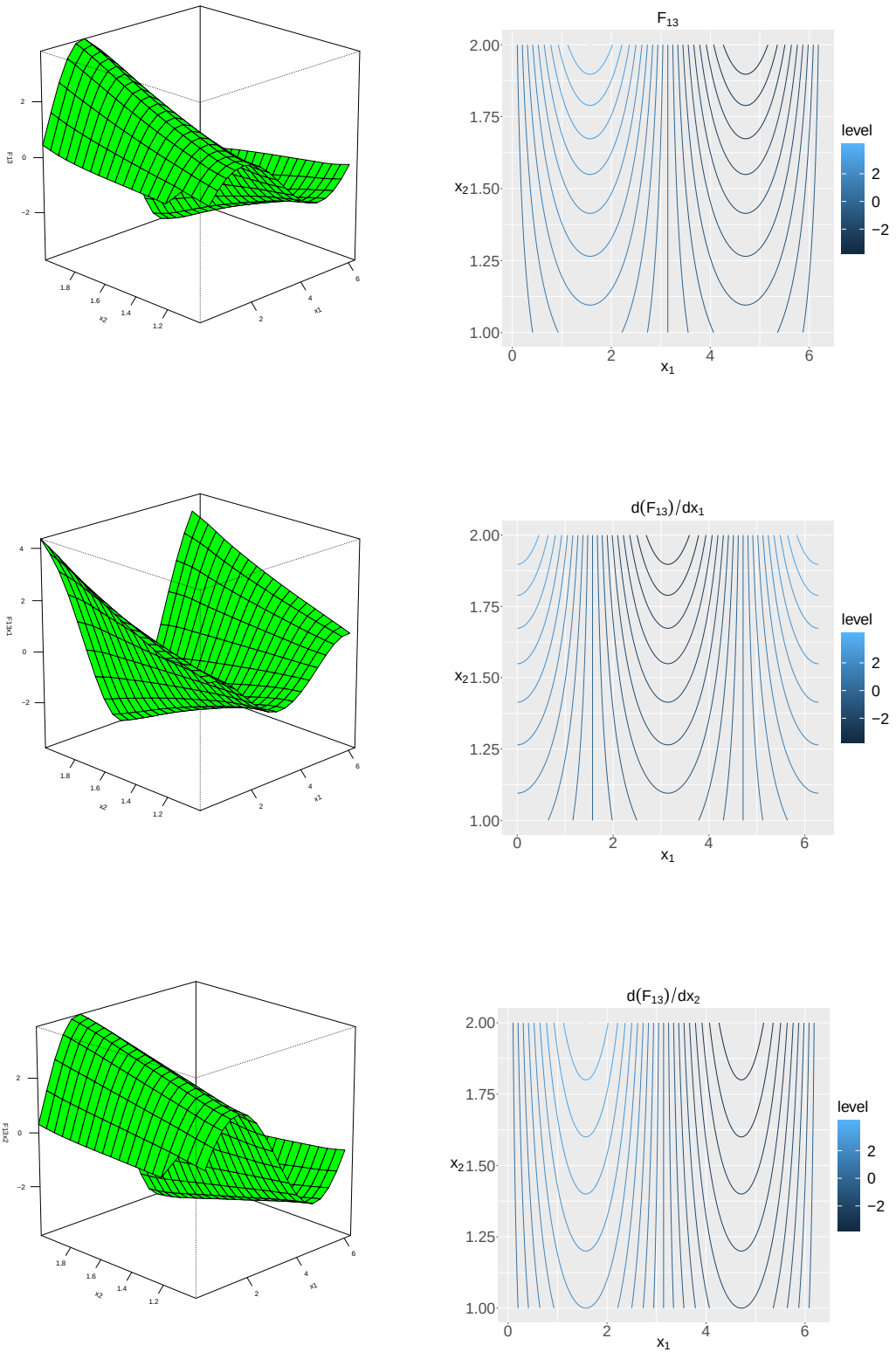


FIGURE 6.13: Perspective and contour plots of F_{13}, $\frac{\partial F_{13}}{\partial x_1}$, and $\frac{\partial F_{13}}{\partial x_2}$.

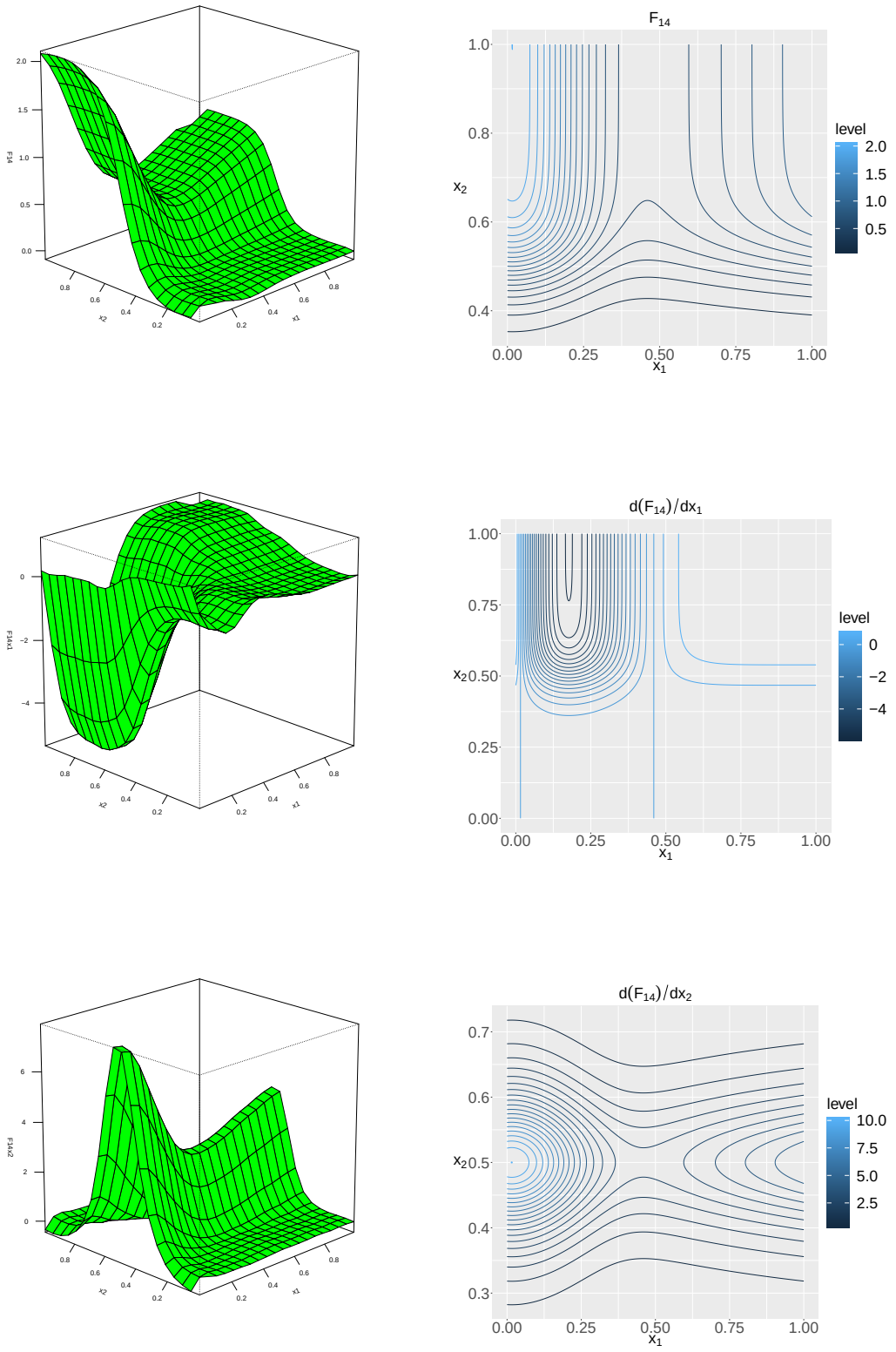


FIGURE 6.14: Perspective and contour plots of F_{14}, $\frac{\partial F_{14}}{\partial x_1}$, and $\frac{\partial F_{14}}{\partial x_2}$.

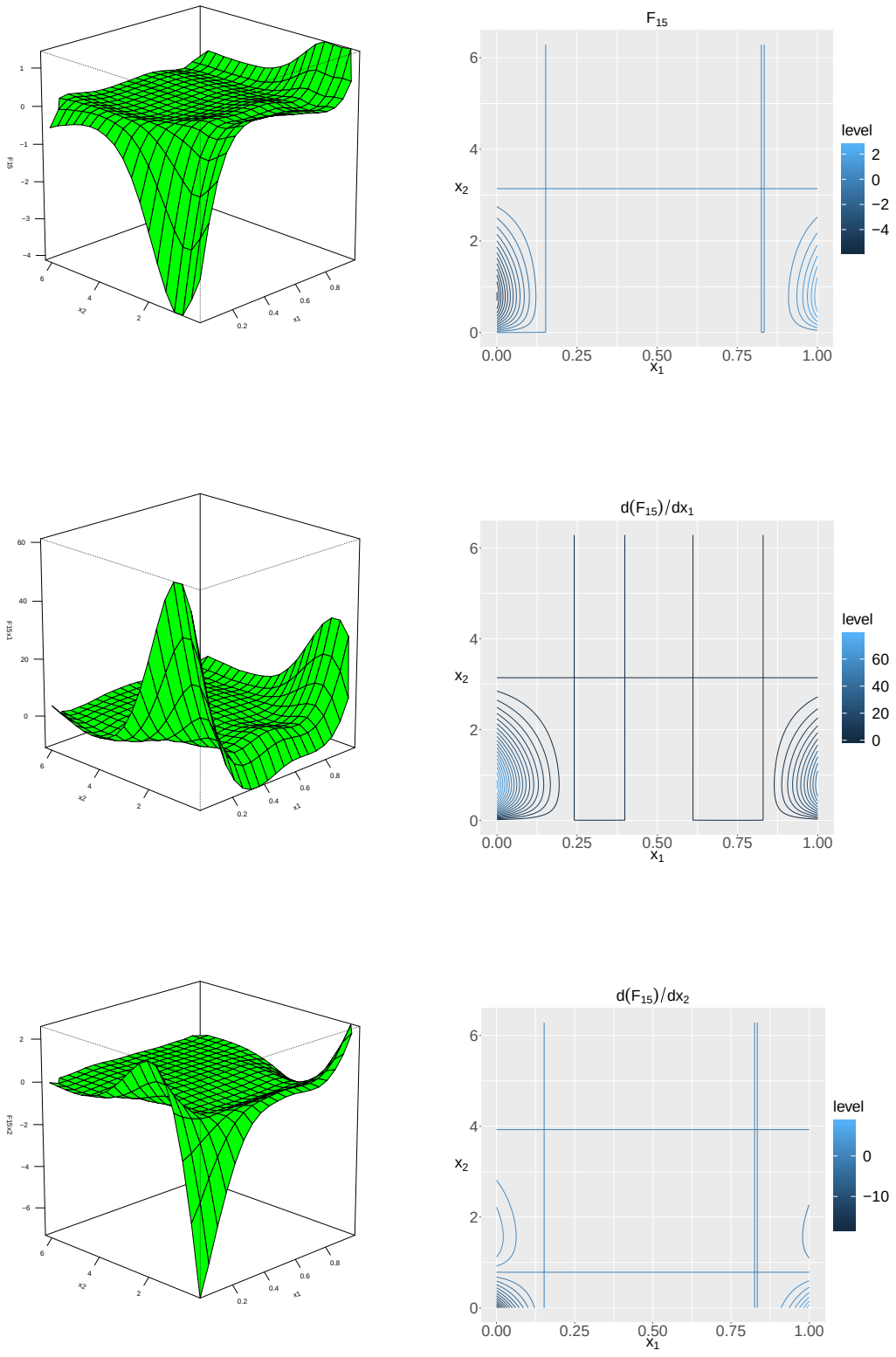


FIGURE 6.15: Perspective and contour plots of F_{15}, $\frac{\partial F_{15}}{\partial x_1}$, and $\frac{\partial F_{15}}{\partial x_2}$.

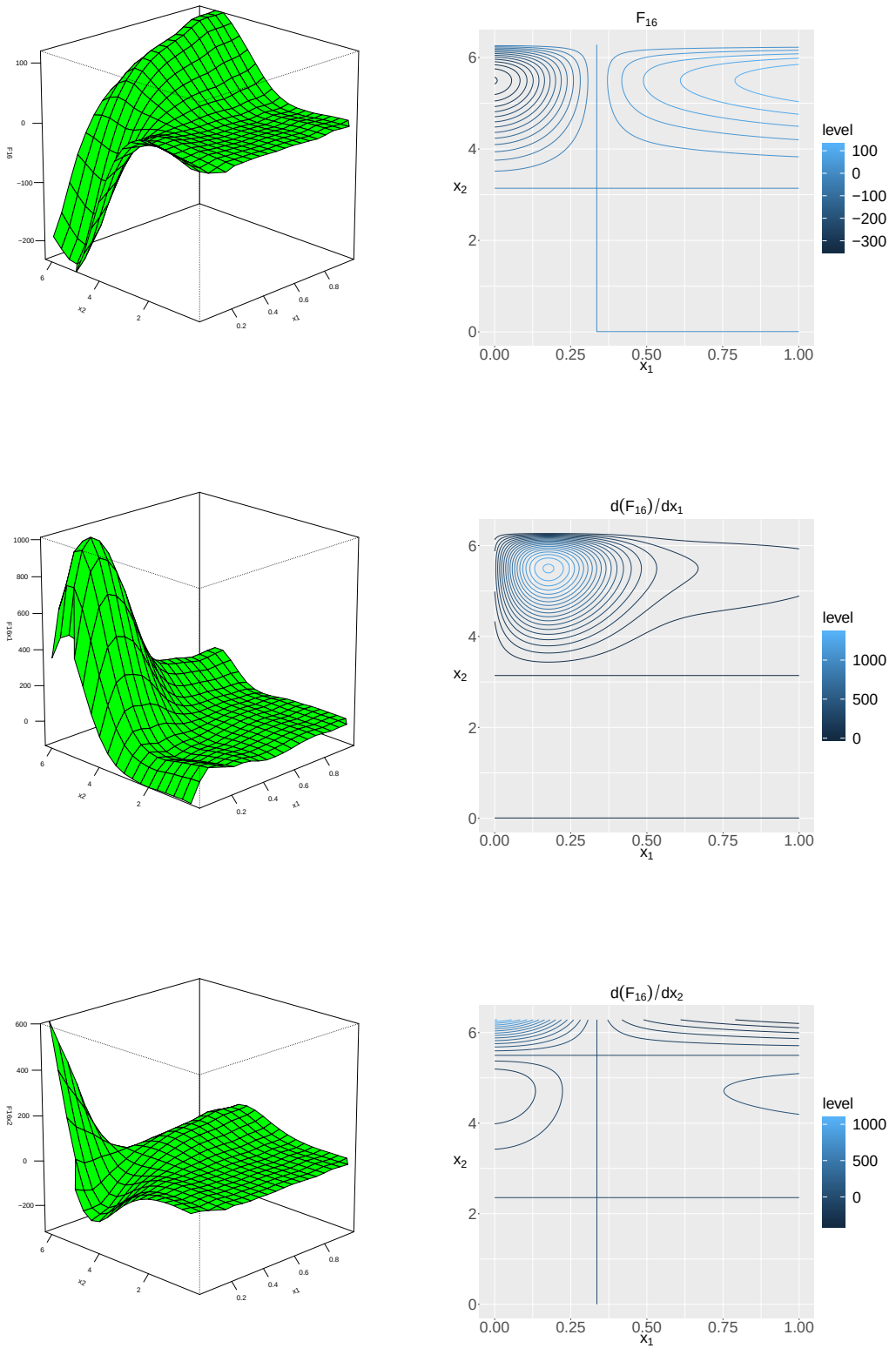


FIGURE 6.16: Perspective and contour plots of F_{16}, $\frac{\partial F_{16}}{\partial x_1}$, and $\frac{\partial F_{16}}{\partial x_2}$.

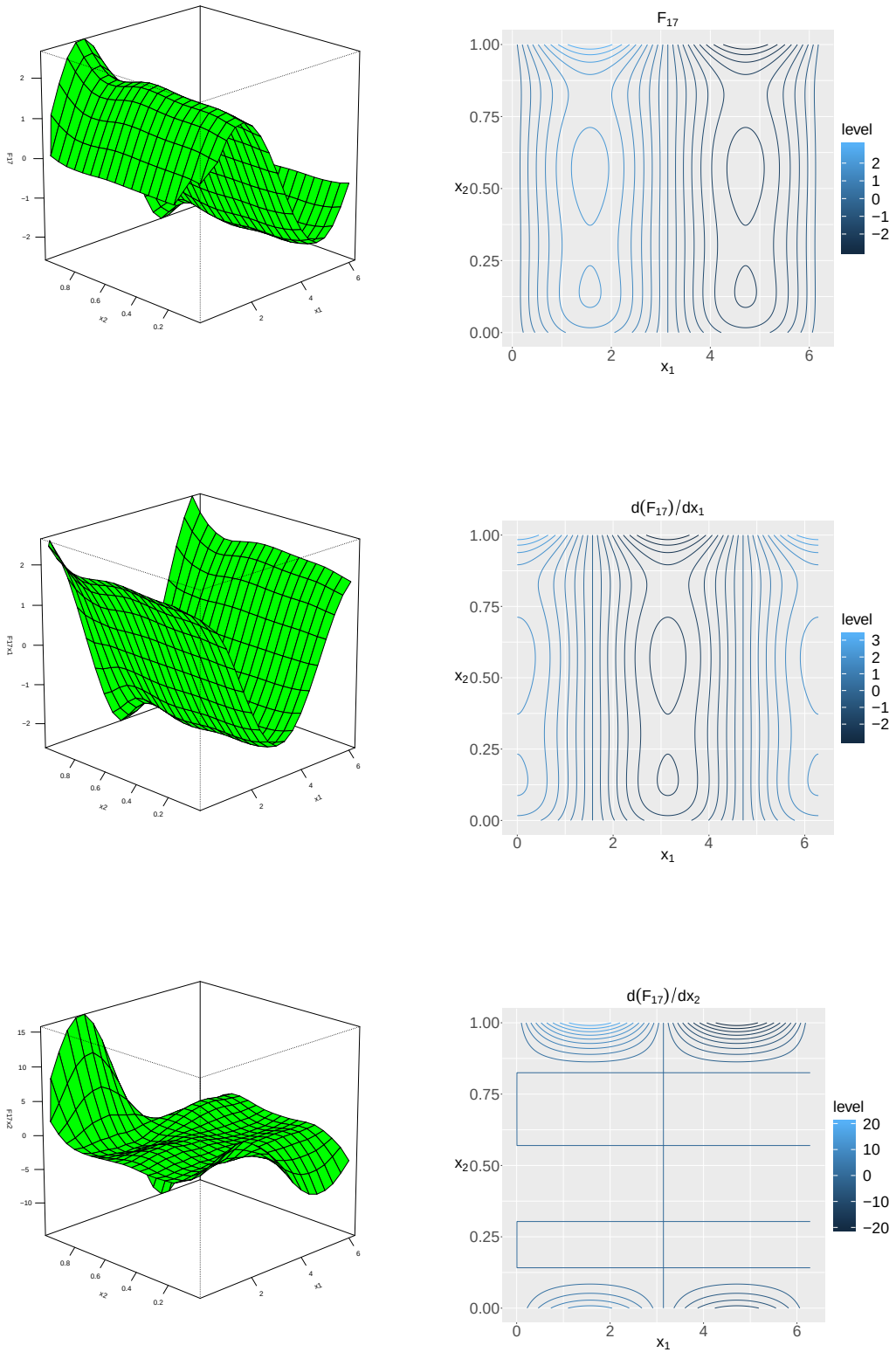


FIGURE 6.17: Perspective and contour plots of F_{17}, $\frac{\partial F_{17}}{\partial x_1}$, and $\frac{\partial F_{17}}{\partial x_2}$.

6.3.2 Simulation Results

This simulation study involves multiple functions with a wide variety of behaviors ranging from simple to very erratic. Moreover, there are 6 levels of random noise added and 2 sample sizes implemented. The scenarios are too numerous for all of them to be presented in this section. For this reason, only the smallest noise level, $\sigma = 0.02$, are presented here, with all other noise levels presented in Appendix A. Figures 6.18-6.25 depict box-plots of the RMSE for 500 simulations across the different penalty orders and across multiple degrees of freedom within each penalty order . There is also included, to the right of each penalty order grouping, a box-plot of the RMSE of the model using the optimal degrees of freedom as selected by the BIC score. These box-plots make it easy to assess the performance of each penalty order when derivative estimation is the main goal. A small median RMSE suggests a better estimate of the derivative. Please note the scales for the left and right panels are not necessarily the same.

The penalty order performance results shown in all the the box-plots are summarized in Tables A.1-A.4. Each coloured entry represents the penalty order which has the lowest median RMSE for the BIC chosen degrees of freedom, the right most box-plot in the degrees of freedom groupings. The different colours for each penalty order (p1, p12, p2 , p13, p23, and p3) aid in assessing the overall performance of each penalty order at a glance.

The bar-plots in Figures 6.26-6.29 summarize the information given in Tables A.1-A.4. Each bar-plot depicts the frequency each penalty attains the minimum median RMSE for the BIC box-plot. The top panels have each bar partitioned by random noise level (StDev) and the bottom panel by function. The overall results can be summarized as follows:

- Penalty order 3 clearly outperforms all other penalty orders for the univariate functions for 500 data points and even more so for 5000 data points. It is interesting to see $\boldsymbol{F_2}$, $\boldsymbol{F_3}$, and $\boldsymbol{F_9}$, the somewhat erratic univariate functions, almost never selected 3 as the optimal penalty order.

- The top panels of Figures 6.26-6.29 show that noise levels do not tend to accumulate in any one penalty order. This suggests optimal penalty order does not depend on the level of noise. Alternatively, the bottom panels show the that functions do tend to accumulate in specific penalty orders, suggesting the penalty order does depend on the type of function.

- For the bivariate functions, the penalty orders where penalty order 3 is involved (p13, p23, and p3) clearly outperform penalty orders which do not. This is true

for both 500 and 5000 data points. Penalty order p23 performs the best for 500 data points, with p3 and p13 coming in 2^{nd} and 3^{rd}. Penalty order p3 performs the best for 5000 data points, with p13 and p23 coming in 2^{nd} and 3^{rd}.

- The percentage of times each penalty order is found optimal is presented in Table 6.1.

Penalty Order	Percentage of Times Found Optimal
p1	0.4%
p12	8.3%
p2	6.5%
p13	22.8%
p23	19.6%
p3	42.4%

TABLE 6.1: Percentages each penalty order is found optimal.

Overall, penalty orders involving penalty order 3 outperform the rest, being optimal 234 times out of 276 (84.8%). They are optimal 105 times out of 132 (79.5%) for univariate functions and 129 times out of 144 (89.6%) for bivariate functions.

Although penalty order 3 is the clear winner for estimating derivatives in this simulation study, the box-plots of Figures 6.18-6.25 and Figures A.1-A.23 show that differences between the medians of penalty order 3 and the optimal penalty orders are very small in many cases. Figures 6.30 and 6.31 show box-plots of the relative difference (RD) defined by

$$RD = \frac{\text{median}(RMSE \text{ for p3}) - \text{median}(RMSE \text{ for optimal penalty order})}{\text{median}(RMSE \text{ for optimal penalty order})}. \quad (6.30)$$

One result which is quickly noticeable is the medians of the relative differences of the partial derivative estimates of bivariate functions in Figure 6.31 are generally less than those of the derivative estimates of univariate functions in Figure 6.30. This is reassuring since the interest lies in partial derivative estimation of the River Run model. The difference in medians between relative differences of penalty order p23 and p3 for 5000 data points for partial derivatives is about 2.5%. This is also reassuring since the River Run model has over 7000 points and p13 and p23 come in a close second and third place, respectively, when estimating partial derivatives.

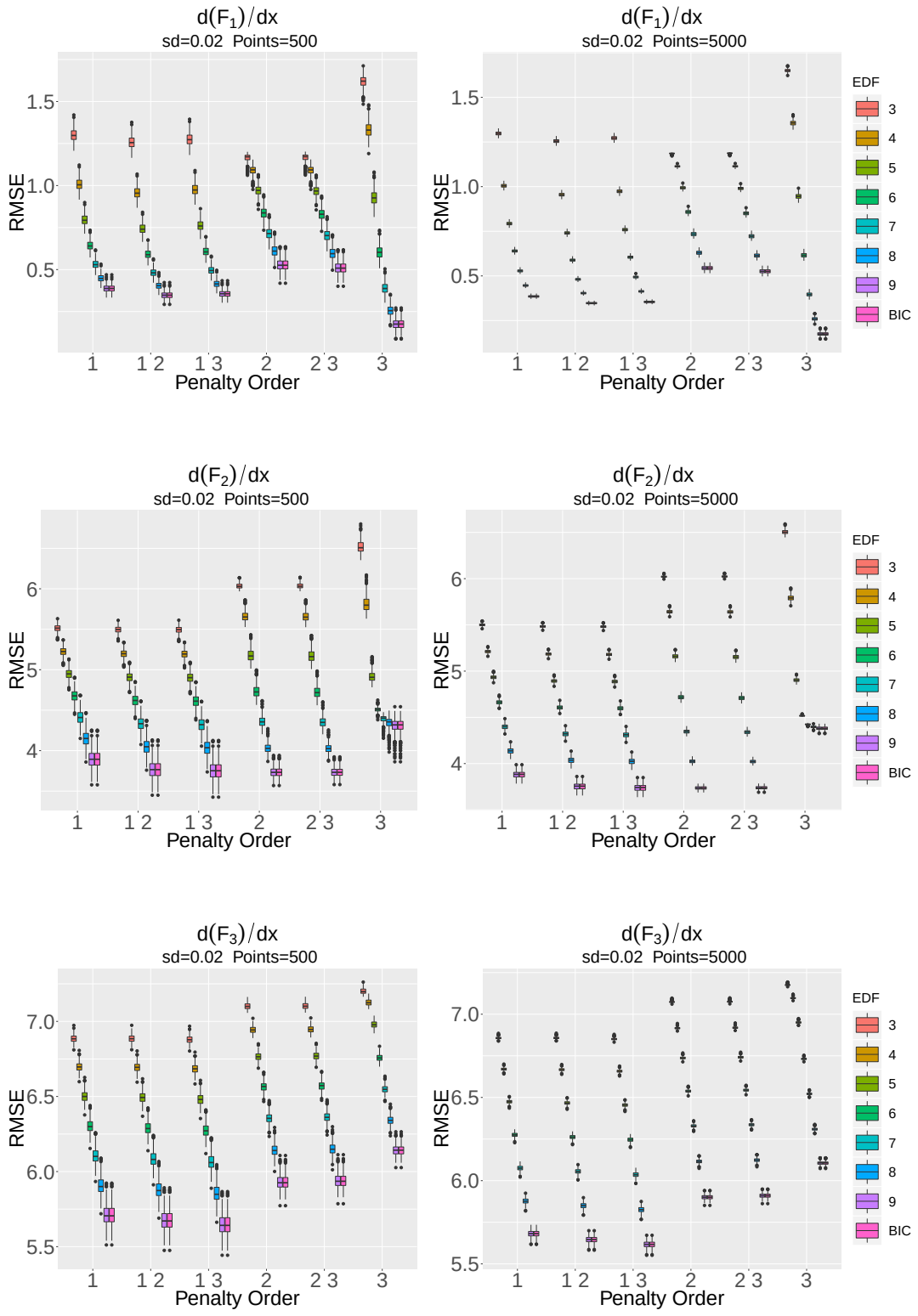


FIGURE 6.18: Box plots of RMSE of $\frac{d(F_1)}{dx}$, $\frac{d(F_2)}{dx}$, and $\frac{d(F_3)}{dx}$ estimates for 500 and 5000 data points.

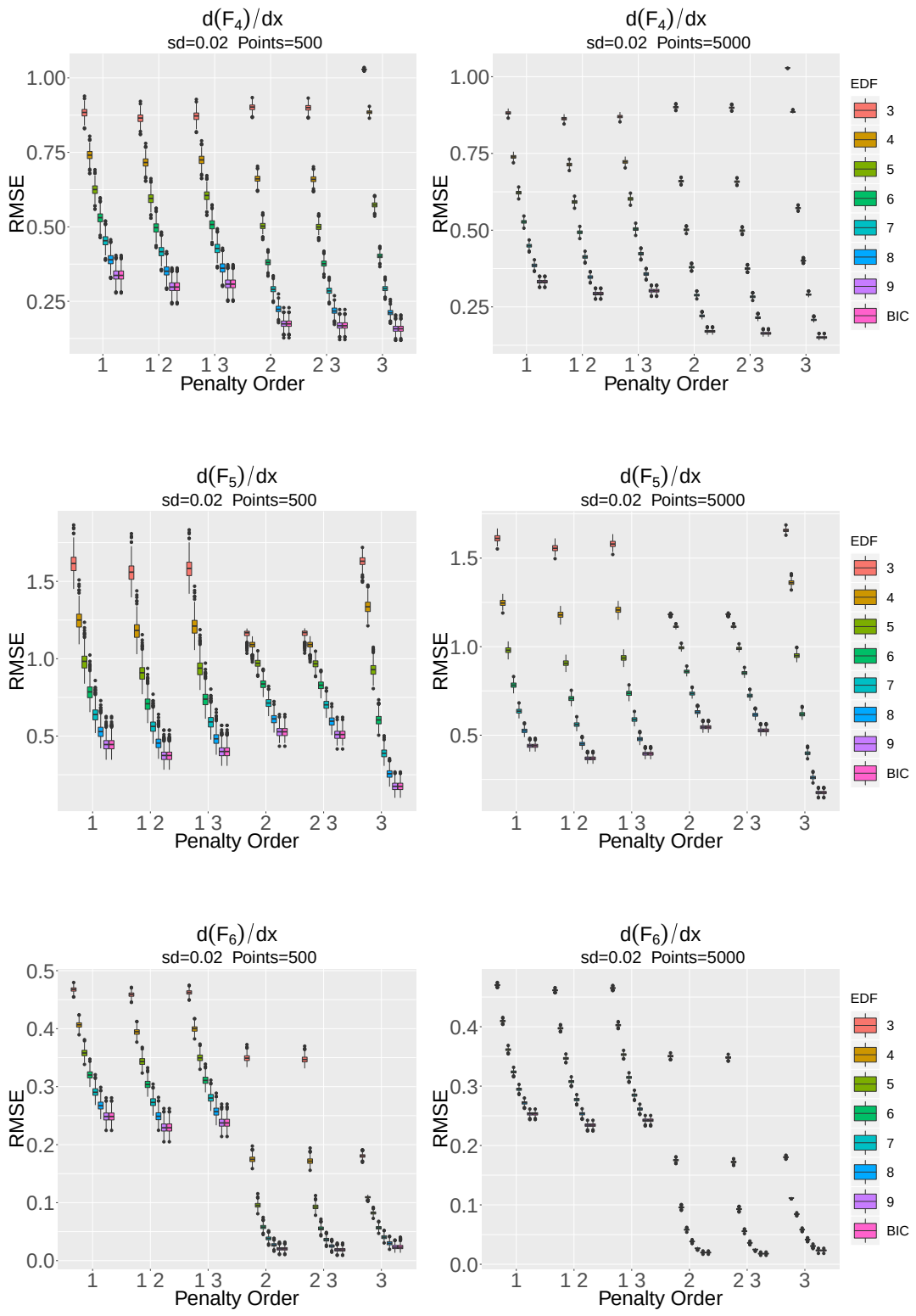


FIGURE 6.19: Box plots of RMSE of $\frac{d(F_4)}{dx}$, $\frac{d(F_5)}{dx}$, and $\frac{d(F_6)}{dx}$ estimates for 500 and 5000 data points.

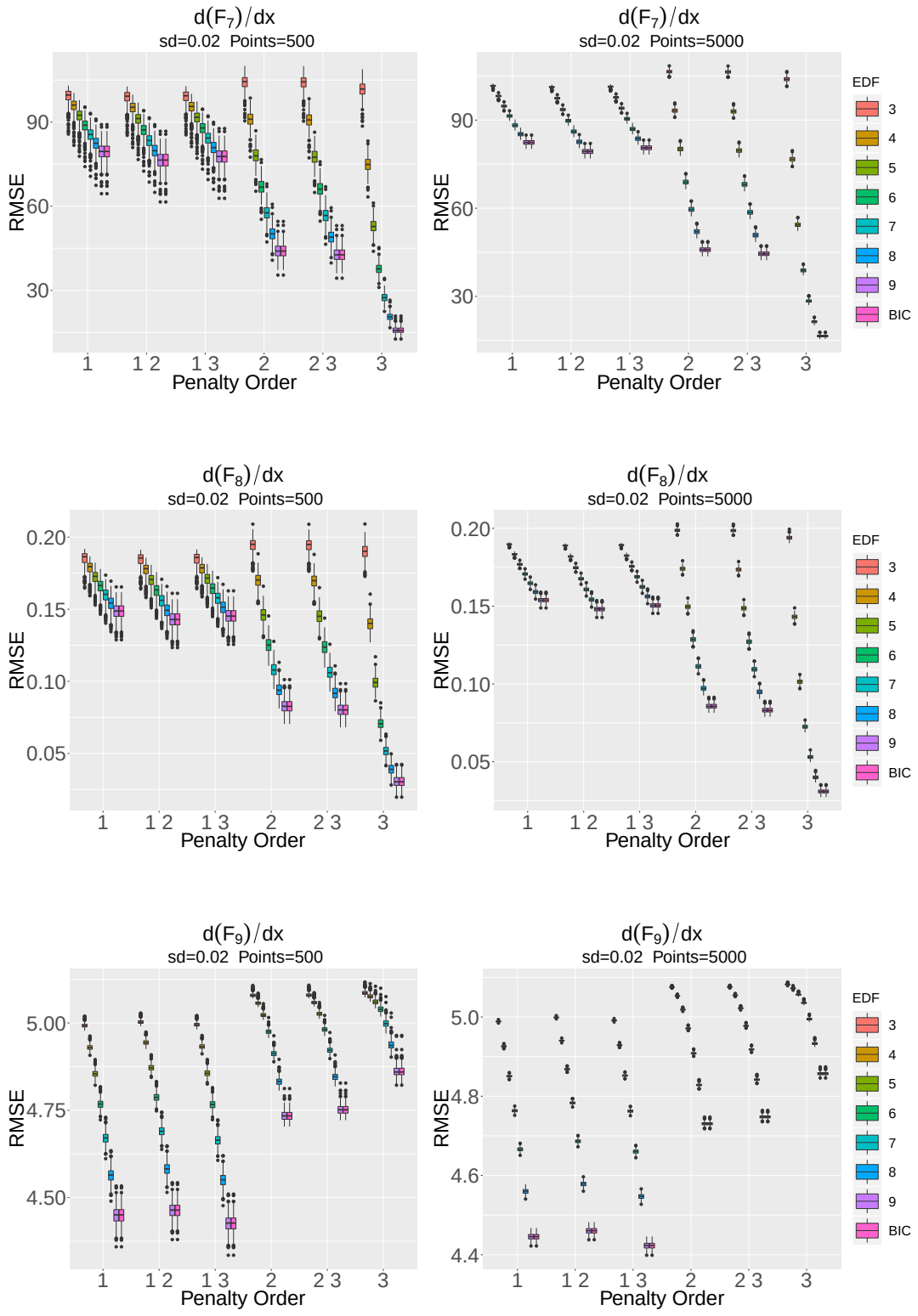


FIGURE 6.20: Box plots of RMSE of $\frac{d(F_7)}{dx}$, $\frac{d(F_8)}{dx}$, and $\frac{d(F_9)}{dx}$ estimates for 500 and 5000 data points.

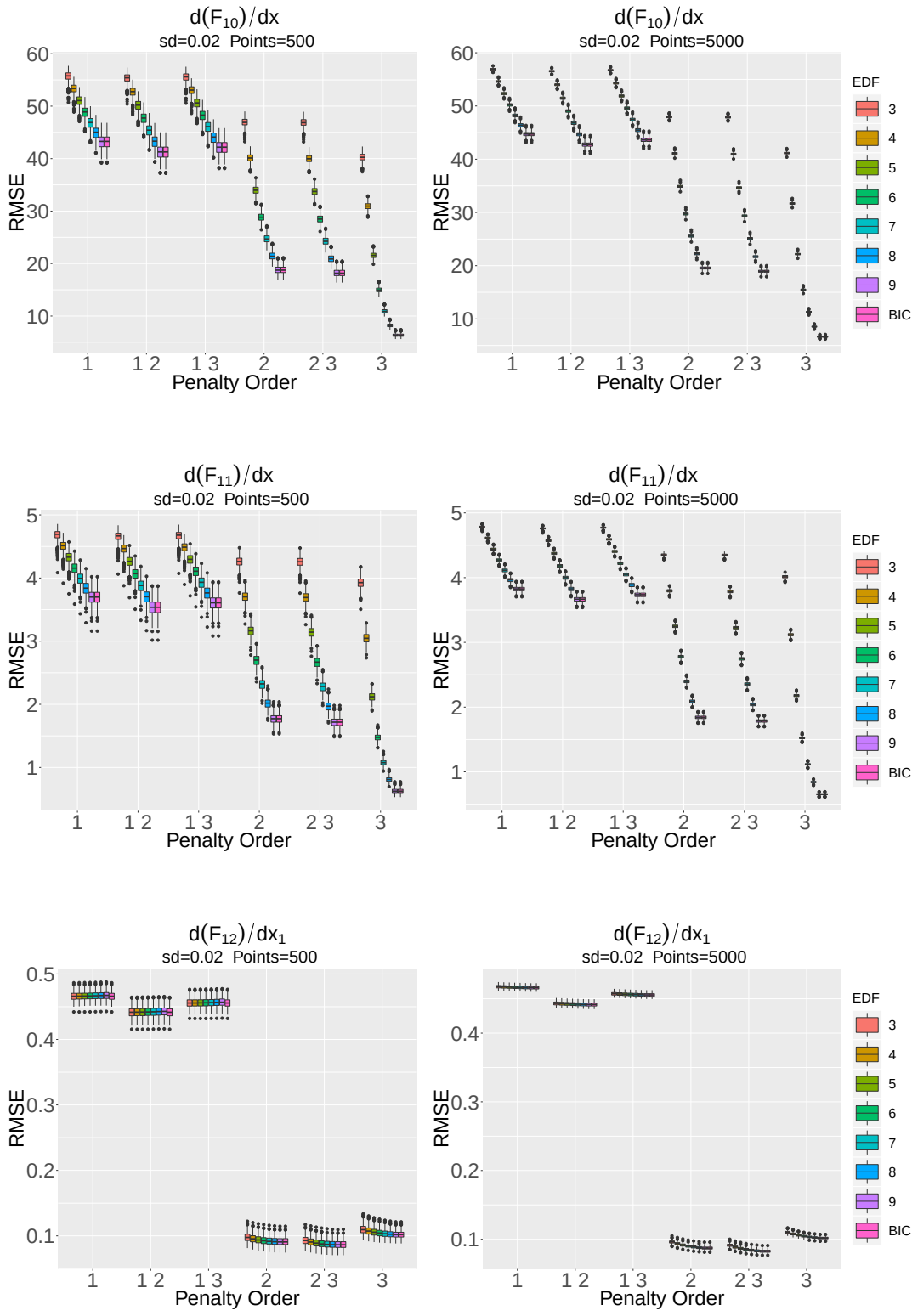


FIGURE 6.21: Box plots of RMSE of $\frac{d(F_{10})}{dx}$, $\frac{d(F_{11})}{dx}$, and $\frac{\partial(F_{12})}{\partial x_1}$ estimates for 500 and 5000 data points.

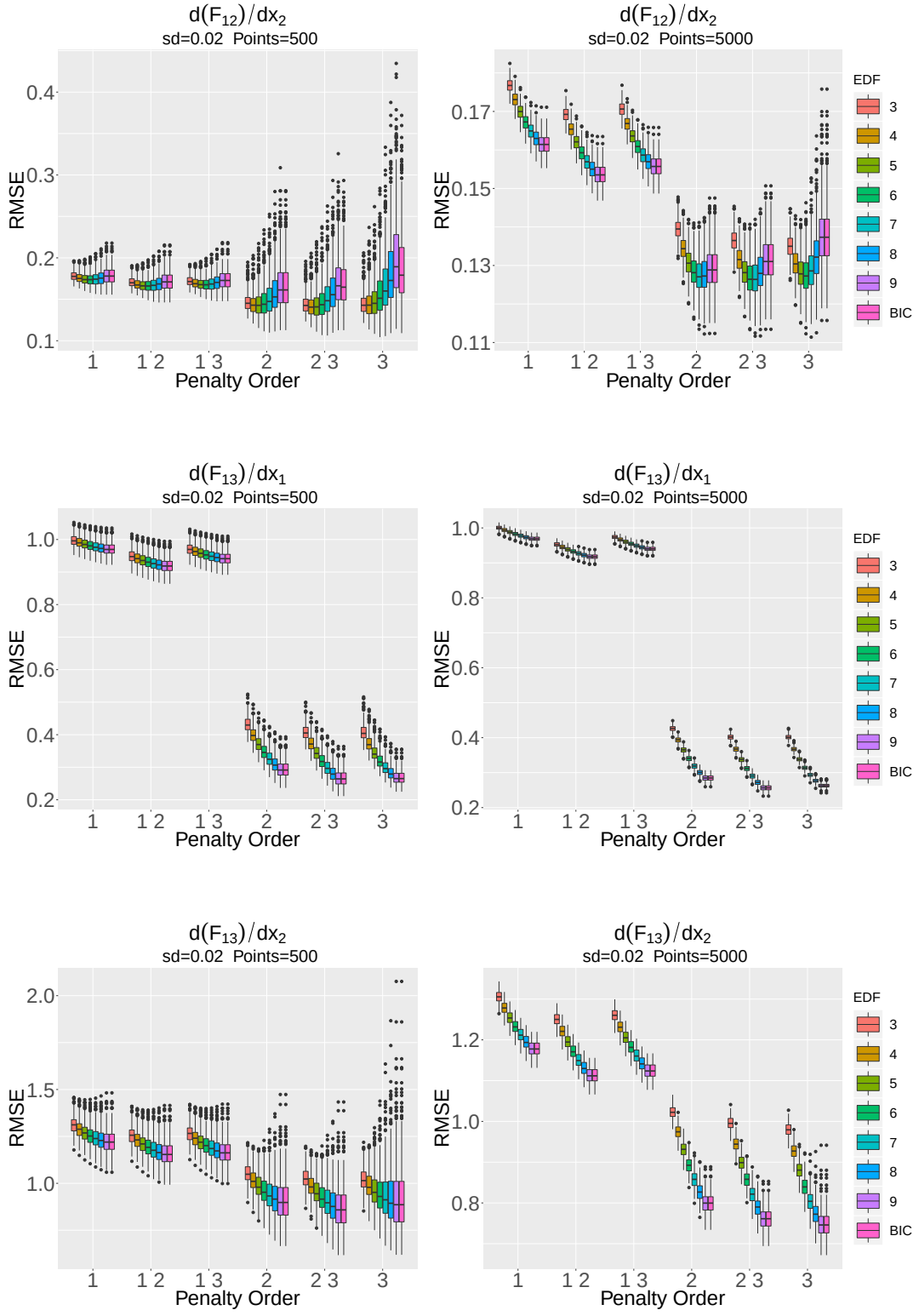


FIGURE 6.22: Box plots of RMSE of $\frac{\partial(F_{12})}{\partial x_2}$, $\frac{\partial(F_{13})}{\partial x_1}$, and $\frac{\partial(F_{13})}{\partial x_2}$ estimates for 500 and 5000 data points.

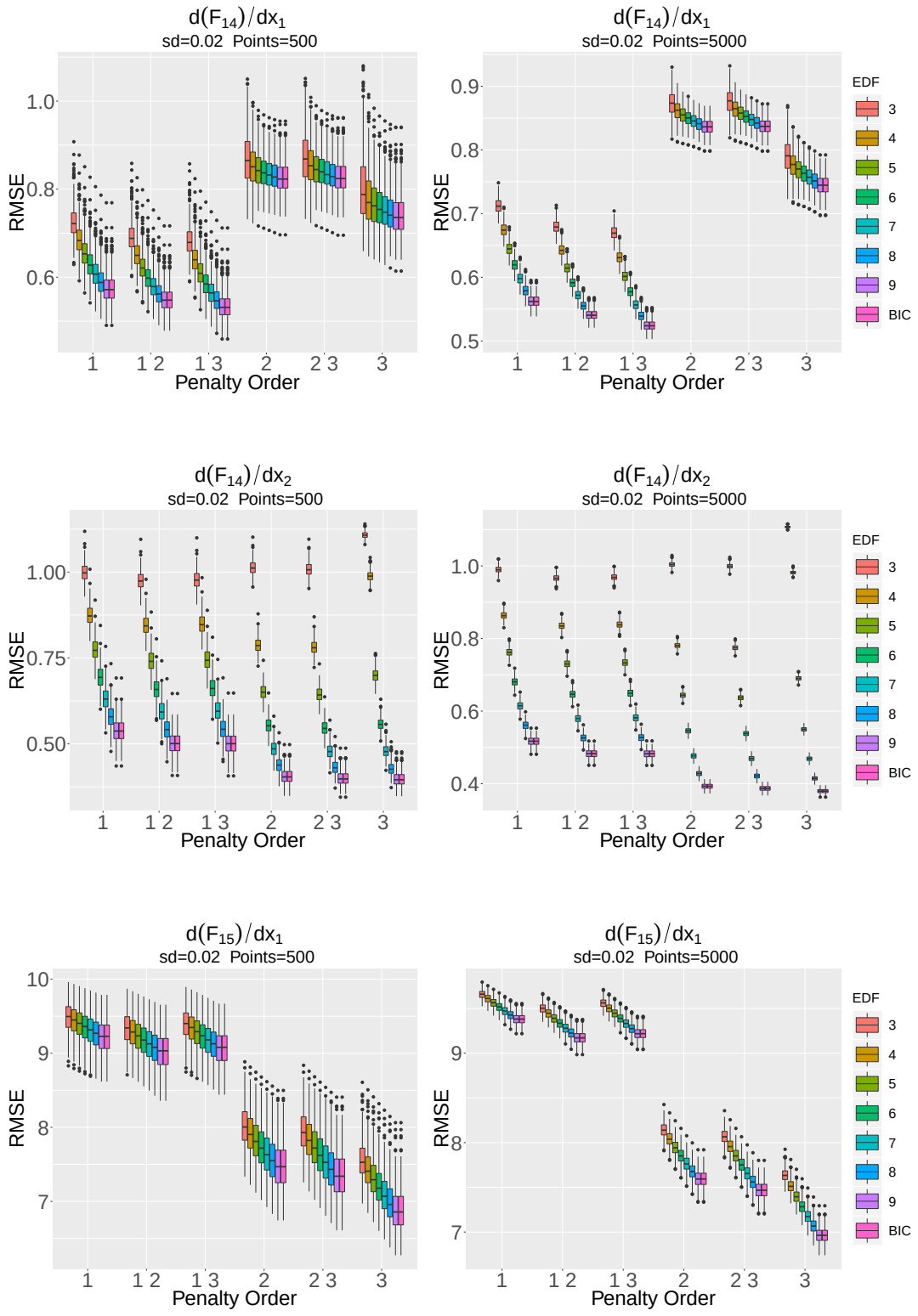


FIGURE 6.23: Box plots of RMSE of $\frac{\partial(F_{14})}{\partial x_1}$, $\frac{\partial(F_{14})}{\partial x_2}$, and $\frac{\partial(F_{15})}{\partial x_1}$ estimates for 500 and 5000 data points.

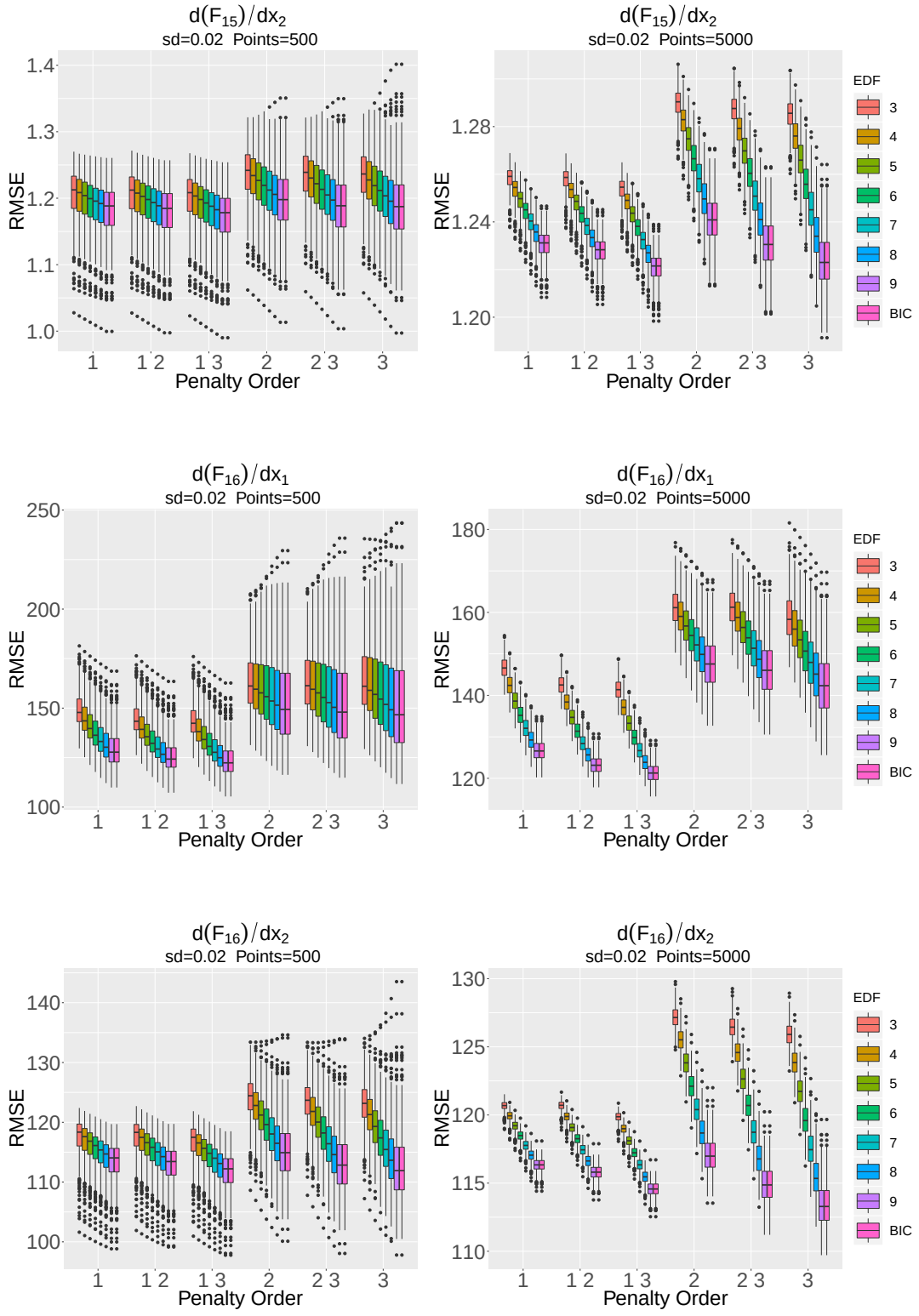


FIGURE 6.24: Box plots of RMSE of $\frac{\partial(F_{15})}{\partial x_2}$, $\frac{\partial(F_{16})}{\partial x_1}$, and $\frac{\partial(F_{16})}{\partial x_2}$ estimates for 500 and 5000 data points.

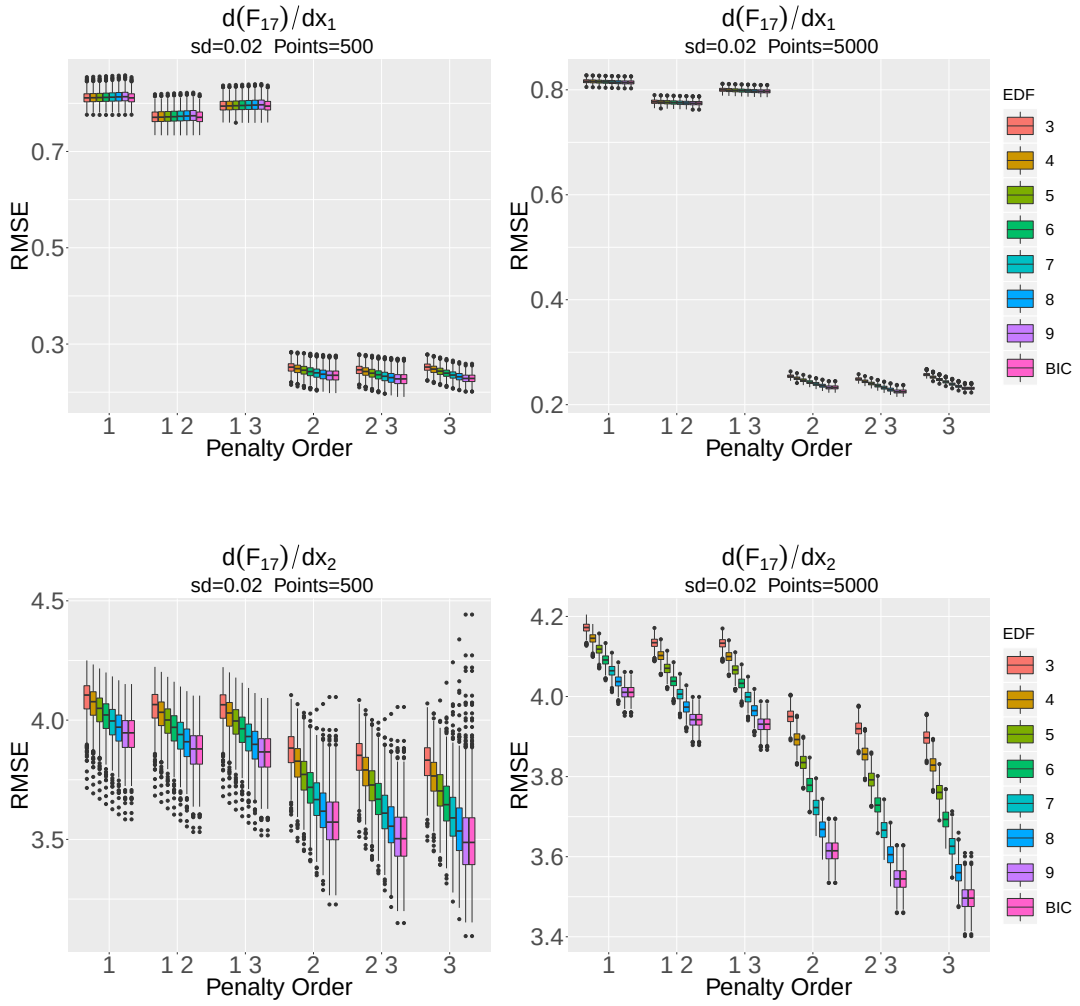


FIGURE 6.25: Box plots of RMSE of $\frac{\partial(F_{17})}{\partial x_1}$, and $\frac{\partial(F_{17})}{\partial x_2}$ estimates for 500 and 5000 data points.

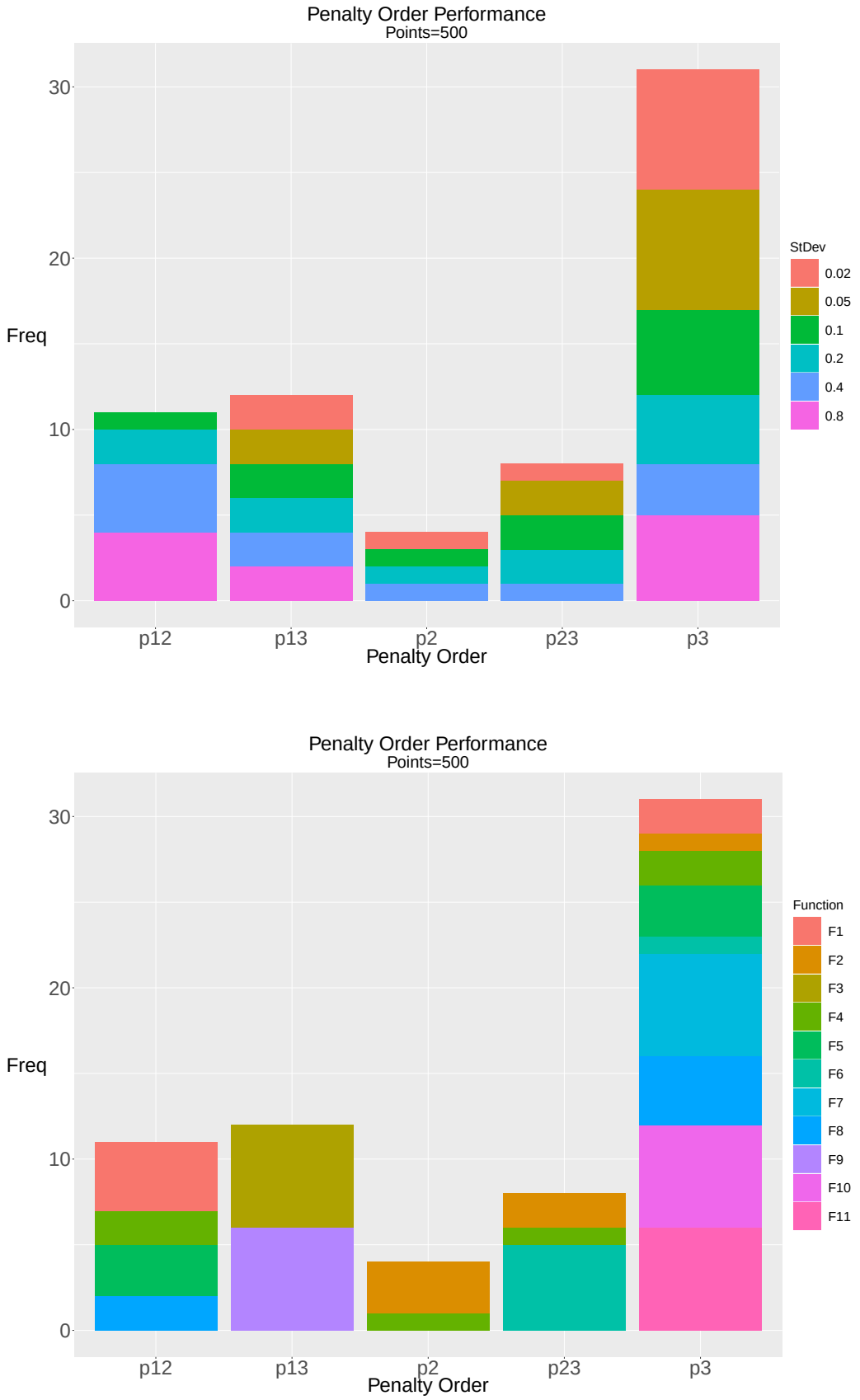


FIGURE 6.26: Plots of penalty order performance when estimating derivatives of univariate functions across noise level and functions for 500 data points.

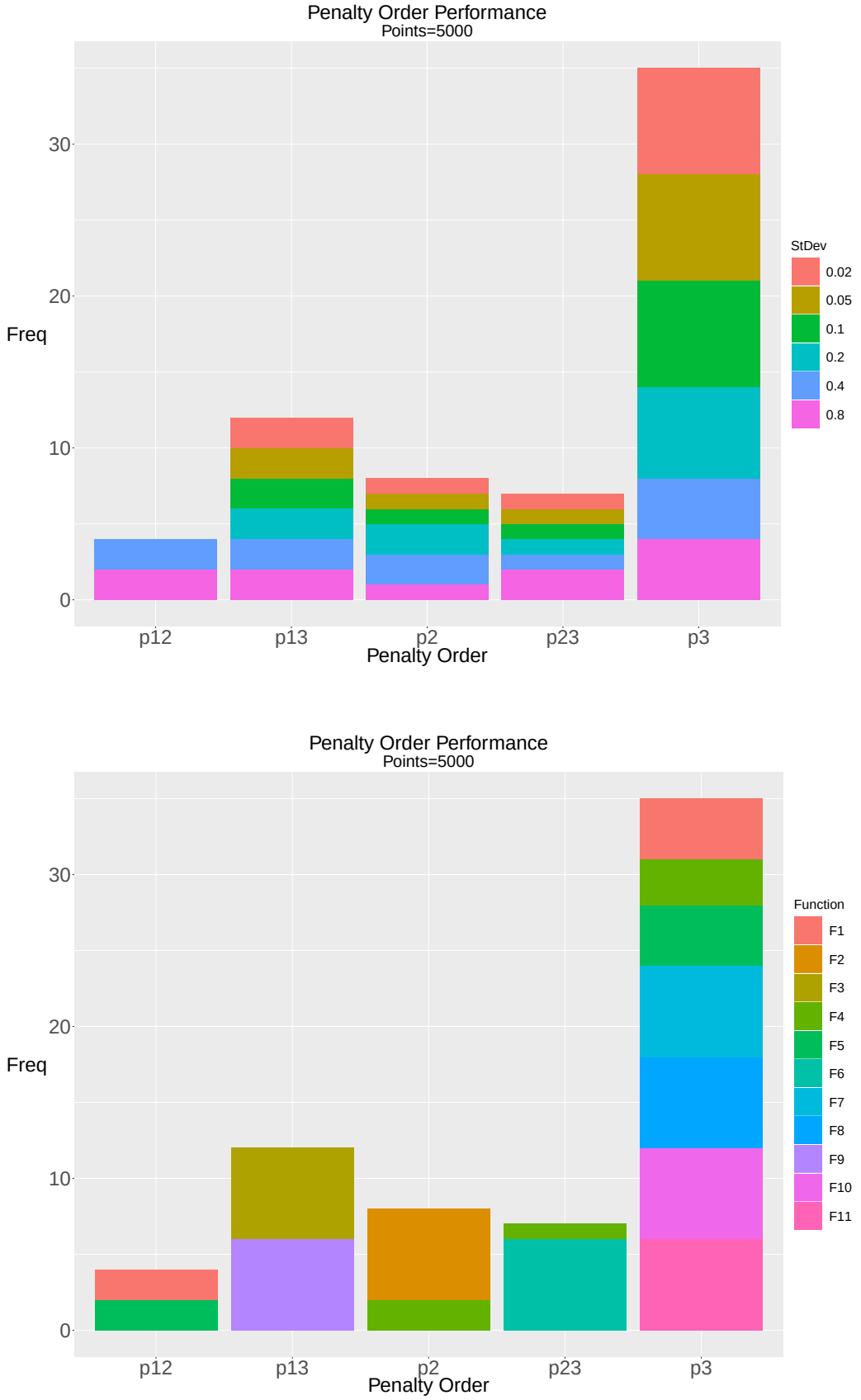


FIGURE 6.27: Plots of penalty order performance when estimating derivatives of univariate functions across noise level and functions for 5000 data points.

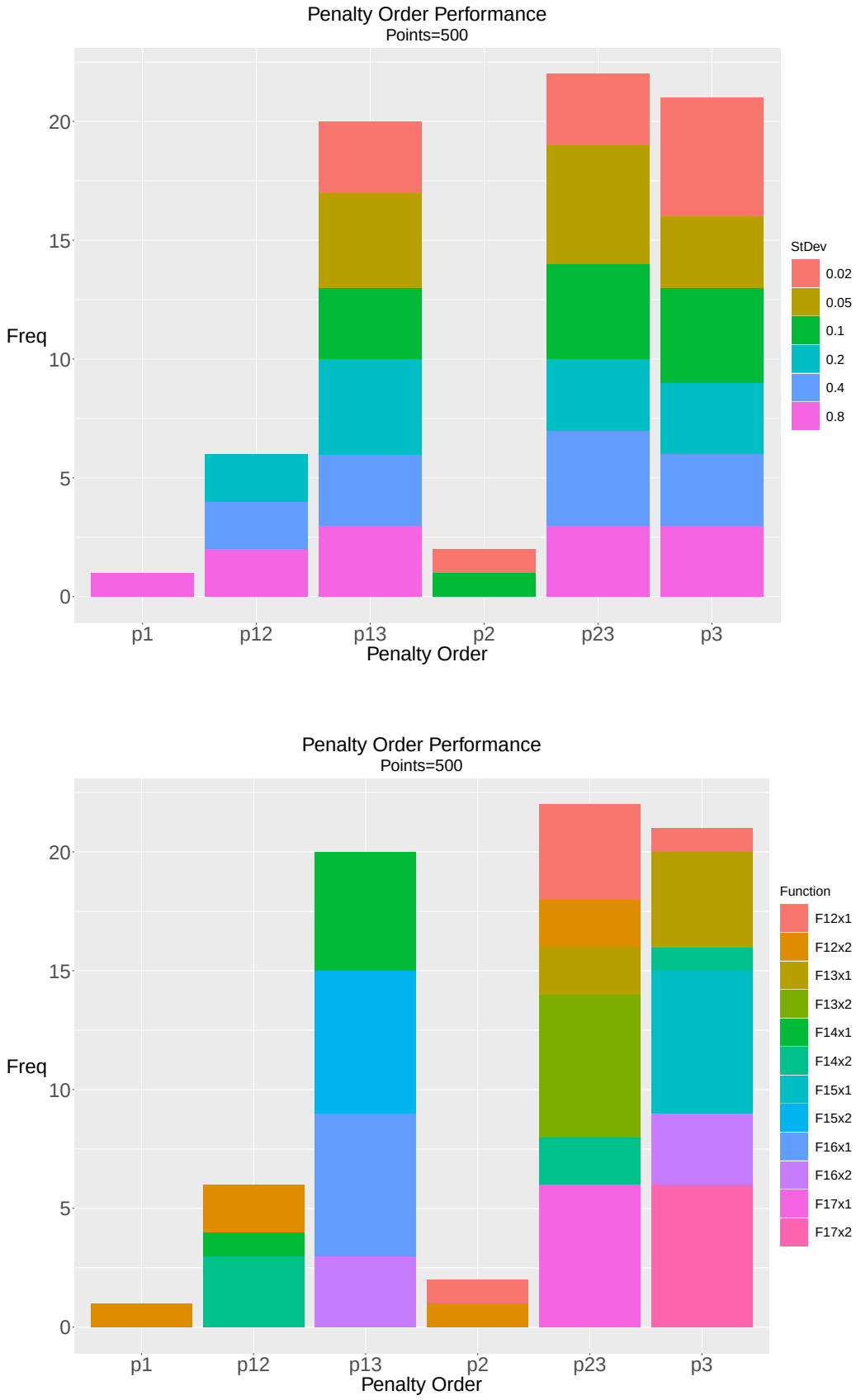
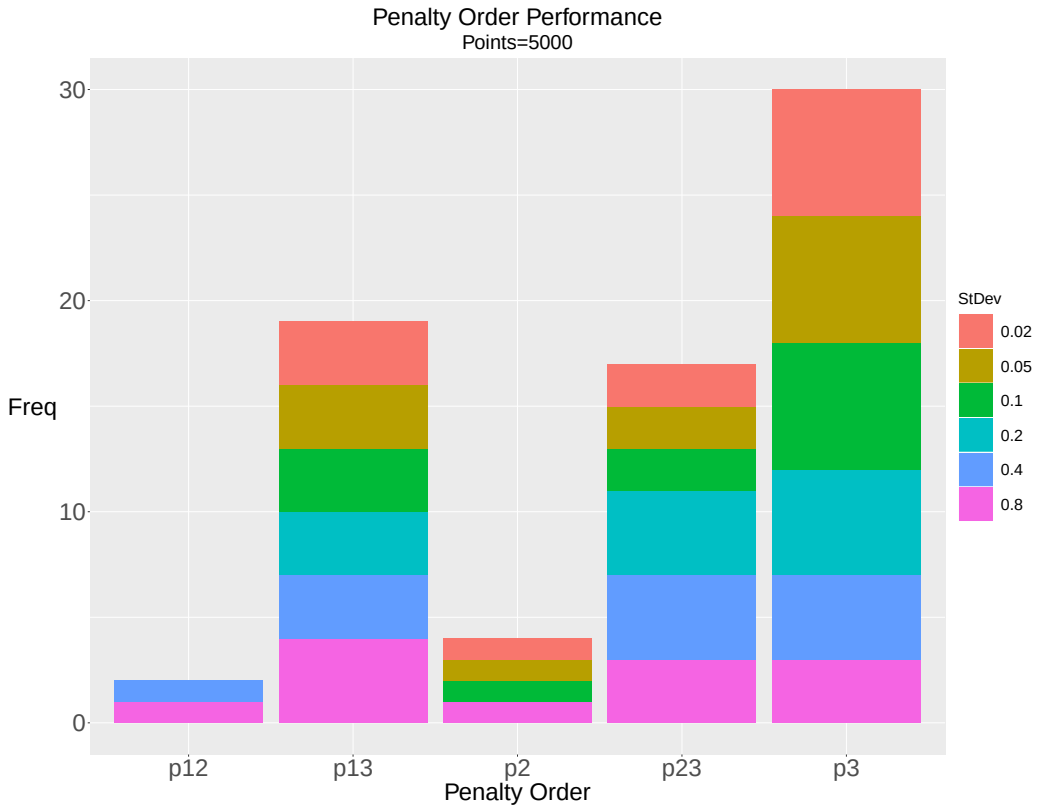


FIGURE 6.28: Plots of penalty order performance when estimating partial derivatives of bivariate functions across noise level and functions for 500 data points.



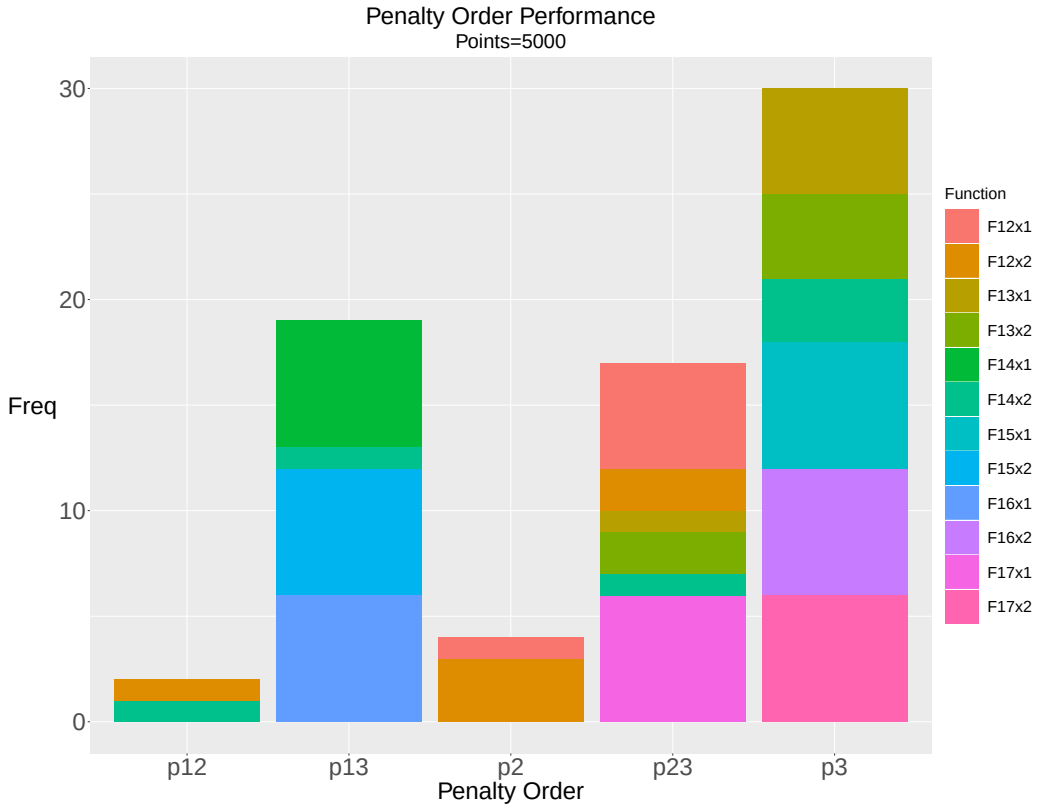


FIGURE 6.29: Plots of penalty order performance when estimating partial derivatives of bivariate functions across noise level and functions for 5000 data points.

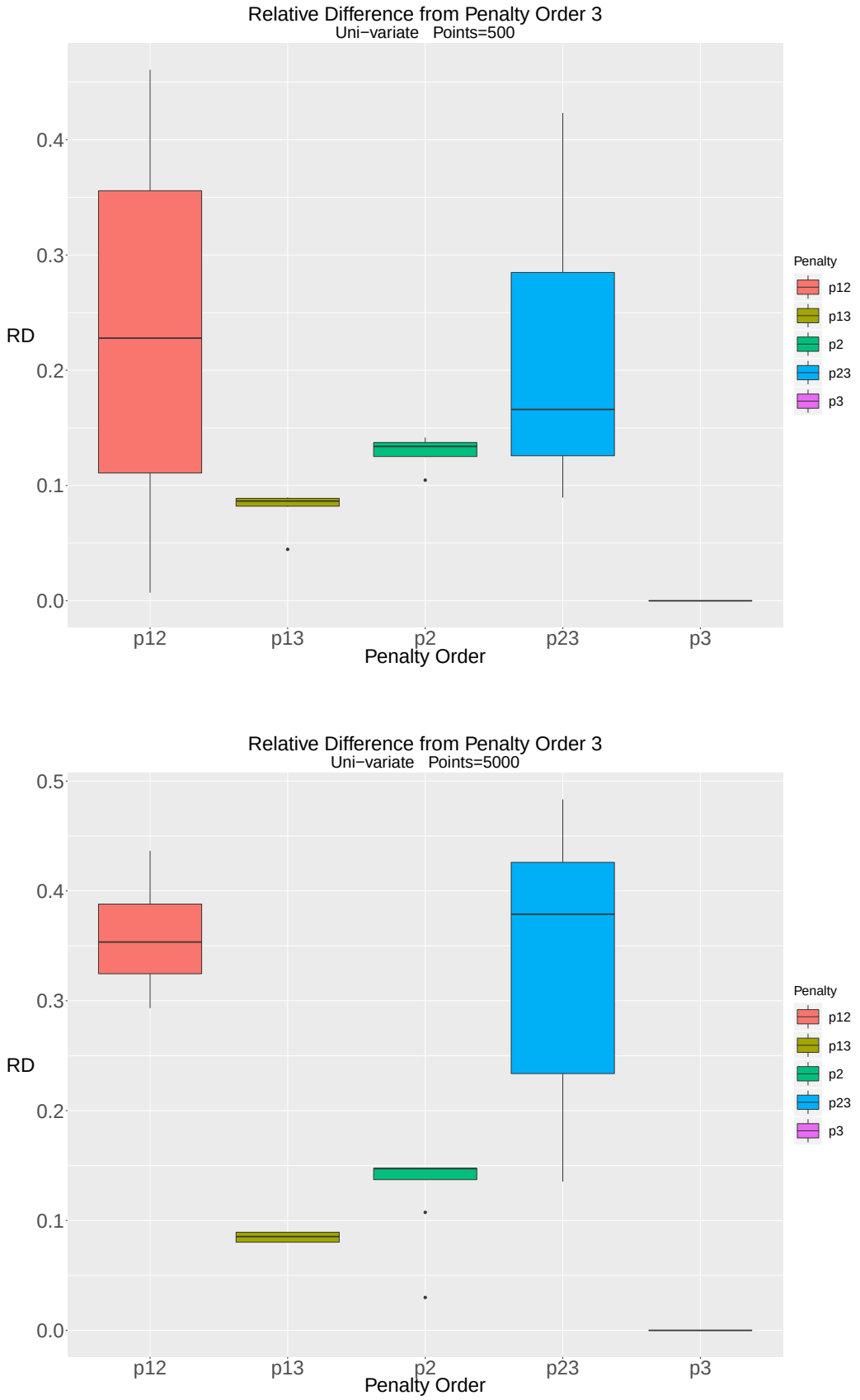


FIGURE 6.30: Box-plots showing the relative difference between penalty order 3 and the optimal penalty order for univariate functions using 500 and 5000 data points.

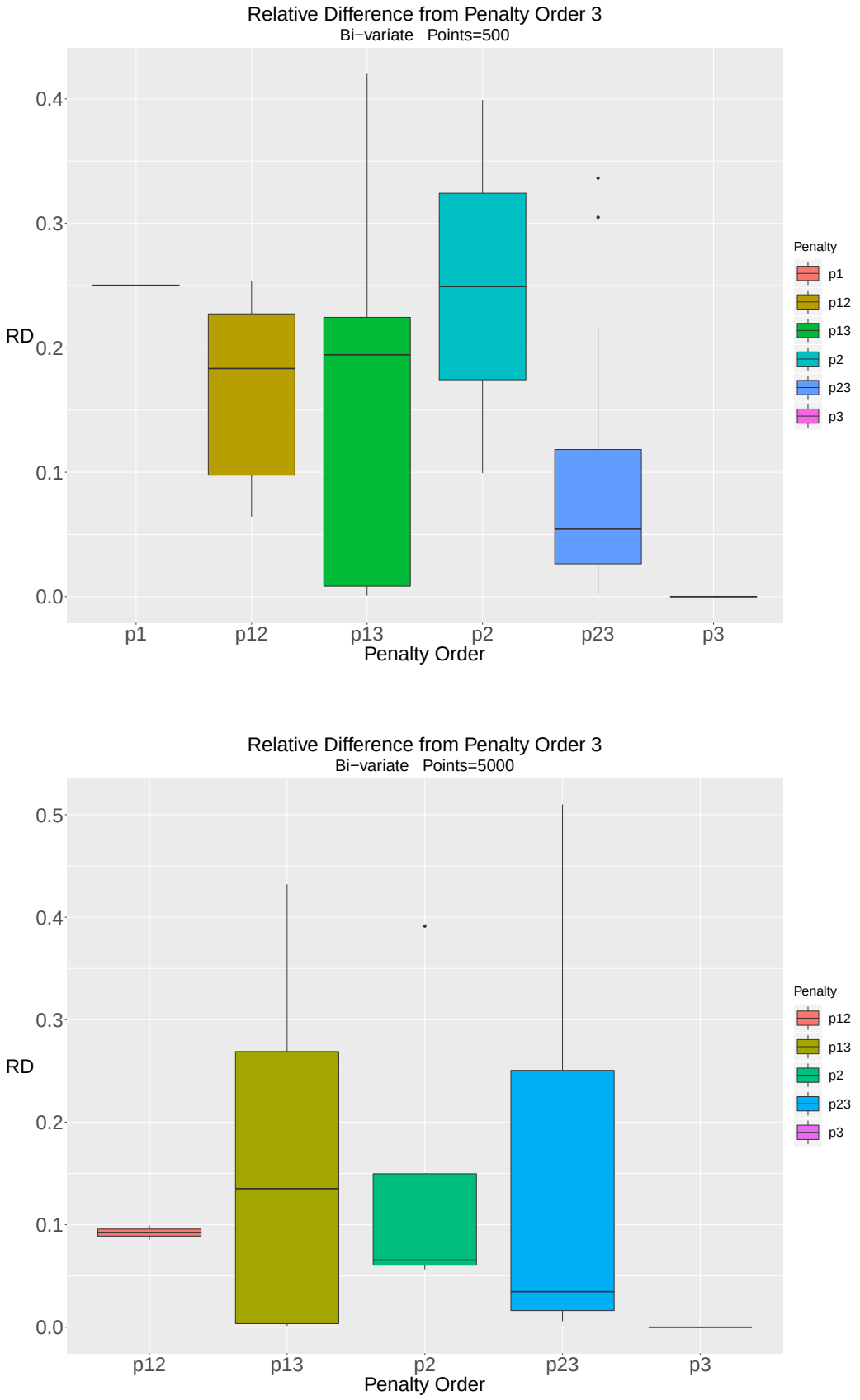


FIGURE 6.31: Box-plots showing the relative difference between penalty order 3 and the optimal penalty order for bivariate functions using 500 and 5000 data points..

6.4 Summary

The choice of penalty order may not be of great importance when constructing a P-spline model since in general there is nothing special about the integrated square of a particular order derivative (Eilers and Marx (1996)). A penalty on the curvature is frequently used, which translates to a second order penalty P-spline model. Here, it is argued that a third order penalty on a P-spline function f translates to a second order penalty on the derivative f'. Such a penalty can be interpreted as a curvature penalty on the derivative f'. The simulation study above suggests there may exist a preferred penalty order when derivative estimation is the primary goal.

The simulation study implements not just a third order penalty, but various single and double penalty orders on a variety of univariate and bivariate functions with known derivatives and partial derivatives. Several levels of noise are also added to the simulated data. The goal is to see if penalty orders other than penalty order 2 perform better when estimating derivatives and partial derivatives. The degrees of freedom are varied for each function across all penalty orders and an optimal degrees of freedom is found by a minimal BIC score. The RMSE is calculated across all penalty orders over 500 simulation for 500 and 5000 data points. The penalty order with the lowest median RMSE is then deemed the optimal penalty order.

The results in Section 6.3.2 confirm the choice penalty order is of importance when derivative estimation is the primary goal. In fact, the customary penalty order of 2 performs quite poorly when estimating derivatives and partial derivatives, with only penalty order 1 performing worse. Penalty order 3 is found to perform (approximately) as well as or better than all other penalty orders for most univariate functions. The exceptions where penalty order 3 performed poorly were for erratic functions $\boldsymbol{F_2}$ (Figure 6.2), $\boldsymbol{F_3}$ (Figure 6.3), and $\boldsymbol{F_9}$ (Figure 6.9). The results for partial derivative estimation also show that penalty order 3 performs (approximately) as well as or better than all other penalty orders for most bivariate functions, the exceptions being $\boldsymbol{F_{14}}$ wrt x_1 (Figure 6.14 middle panel) and $\boldsymbol{F_{16}}$ wrt x_1 (Figure 6.16 middle panel). Each of these exceptions exhibit extreme changes in their partial derivatives at one locale. This may be the reason penalty order 3 performed poorly in these cases, but this is pure speculation.

The River Run data contains more than 7000 entries and the partial derivatives of bivariate and trivariate interactions will be analyzed. The simulation study scenario which most closely corresponds to the River Run data is partial derivative estimation for 5000 data points. This scenario has penalty order 3 outperforming all other penalty orders and penalty orders 1/3 and 2/3 coming in second and third place respectively, collectively accounting for 84.8% of the optimal penalty orders. Moreover, the median

relative difference for these two penalty orders is approximately 2.5%. This is compelling evidence penalty order 3 should be implemented in a derivative model for the River Run data.

The methods for derivative and partial derivative estimation addressed in Chapter 6 will serve as a means of detecting a sudden increase in the dissolved oxygen levels in a system which has seen a steady increase in those levels over decades. A derivative model will be constructed in Chapter 7, where a penalty order of 3 will be implemented on the Year main effect. Through the heightened sensitivity of bivariate and trivariate interaction terms, this derivative model will ideally allow for the detection of subtle increases in the dissolved oxygen due to the upgrades at the Dalmuir and Shieldhall wastewater treatment facilities.

Chapter 7

A Derivative Model for the River Run Data

The construction of a representative additive mixed model of the River Run not only provides insight to the behaviour of various main effects and interactions driving the dissolved oxygen levels, but also facilitates the detection of changes in those dissolved oxygen levels brought on by known events at particular times and locations. The steady increase in dissolved oxygen over the time span of the River Run data makes the detection of these subtle changes difficult using our model as it stands, even with the inclusion of trivariate interaction terms. However, the added sensitivity of partial derivatives with respect to Year is better suited for detecting any sudden yet subtle increases in the dissolved oxygen. The results of Chapter 6 suggest a penalty order of 3 may also boost the performance of derivative and partial derivative estimation, specifically for data sets containing thousands of points. In this chapter model 5.5 will be fitted as it was in Chapter 5 with the exception of a 3^{rd} order penalty being applied to the Year term. This will be called the derivative mixed model or the derivative model.

Upgrades of the wastewater treatment facilities at Shieldhall in 1985 and Dalmuir in 2001-2004 are the specific known events this analysis is attempting to detect. Once the model is fitted, plots of the partial derivative with respect to Year of the Day of Year : Station : Year interaction with lower terms included will be presented with the focus on the location and timimg of the upgrades.

7.1 Partial Derivative Estimation - The Derivative Model

The same method of minimum BIC score is used to select the degrees of freedom in the penalty order 3 model. The selected degrees of freedom are shown in Table 7.1. There are increased degrees of freedom for Day of Year, Station, Year, Temperature, and River Flow as compared to these given in Table 5.2 for a similar model with penalty order 2.

Main Effect	Degrees of Freedom - BIC
Day of Year	5
Station	3
Year	5
Temperature	6
Salinity	6
Tide	4
Spring	4
River Flow	3

TABLE 7.1: Table of main effects and corresponding degrees of freedom as chosen by BIC for complex additive mixed model with penalty order 3.

7.1.1 Main Effects and Bivariate Interaction Terms

The main effects and interactions of the complex additive mixed model with penalty order 3 for Year fitted here will be depicted as they were in Chapter 5 with the exception of terms involving Year. The derivative and partial derivative with respect to Year will be depicted in these cases. The reason for this is the penalty order 3 was implemented specifically for a derivative model for the River Run data. These derivatives and partial derivatives are calculated using equation 6.9.

Figure 7.1 depicts the main effects. With the exception of Year main effect derivative, all other main effects are very similar to penalty order 2 model. There are very slight increases in flexibility for the main effects where the degrees of freedom increased. Panel c) depicting the derivative of the Year main effect with respect to Year does achieve a local maximum just past the Year 2000 which is just prior to the outset of the upgrades to Dalmuir in 2001, but fails to detect the Shielhall upgrades in 1985. Of course, this main effect involves the whole river and does not pinpoint specific Stations.

The bivariate interaction plots of Figures 7.2-7.4 tell a more interesting story. Although the plots not involving Year are very similar to the corresponding plots of the penalty order 2 complex additive mixed model, some of the plots involving Year seem to have a

common theme; something interesting is happening 1985 and (to a lesser extent) 2001. The behaviour of the Year interactions can be summarized as follows:

- Day of Year : Year (Figure 7.2b) — Negative adjustments of at least -2 standard errors are present circa 1985 autumn months. Positive adjustments of at least +3 standard errors are present circa 1985 for the spring months and of at least +2 circa 2003 for the summer months.
- Temperature : Year (Figure 7.2d) — Negative adjustments of at least -3 standard errors are present circa 1985 for low temperatures and circa 1970 for middle temperatures. Positive adjustments of at least +4 standard errors are present circa 1988 for middle temperatures and circa 1970 for high temperatures.
- Salinity : Year (Figure 7.2f) — Negative adjustments of at least -8 standard errors are present circa 2000 for a salinity of about 5. Positive adjustments of at least +7 standard errors are present circa 1983 for a salinity of about 16 and of at least +9 circa 2000 for a salinity of 21.
- Spring : Year (Figure 7.3b) — Negative adjustments of at least -2 standard errors are not present. Positive adjustments of at least +2 standard errors are present circa 1985 at approximately maximum Spring tide.
- Tide : Year (Figure 7.3d) — Negative adjustments of at least -7 standard errors are present circa 1982 at high Tide and circa 2012 just prior to high Tide. Positive adjustments of at least +6 standard errors are present circa 1980 low tide and of at least +4 circa 2003 just prior to high Tide and circa 2016 just after high Tide .
- River Flow : Year (Figure 7.3f) — Negative adjustments of at least -2 standard errors are present circa 1995 for a River Flow of 170. Positive adjustments of at least +2 standard errors are present circa 1975 for a River Flow of 70 and circa 2003 for a River Flow of 10.
- Station : Year (Figure 7.4 — Negative adjustments getting a low as -7 standard errors are present in a time period ranging from 1970 to 2016 for Stations 14 to 24. Positive adjustments going as high as +12 standard errors are present in a time period ranging from 1970 to 2016 for Stations 0 to 14.

As was the case for the complex additive mixed model fitted in Chapter 5, the derivative model has bivariate interaction that show considerable adjustment of multiple standard errors. This is strong evidence these interactions should be included in the derivative model.

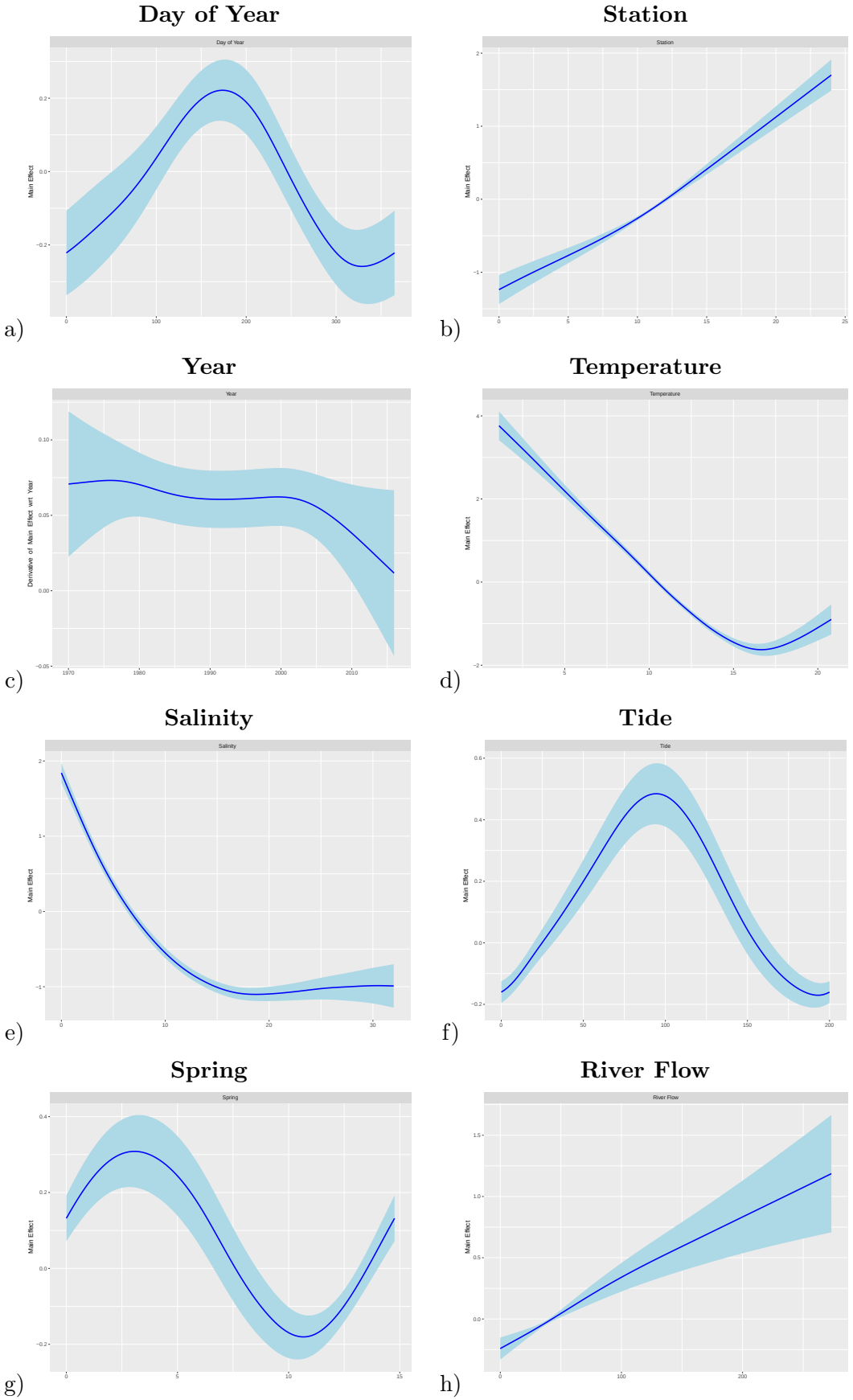


FIGURE 7.1: Plots of main effects a) Day of Year, b) Station, c) derivative Year main effect wrt Year, d) Temperature, e) Salinity, f) Tide, Spring, and g) River Flow on DO for derivative model.

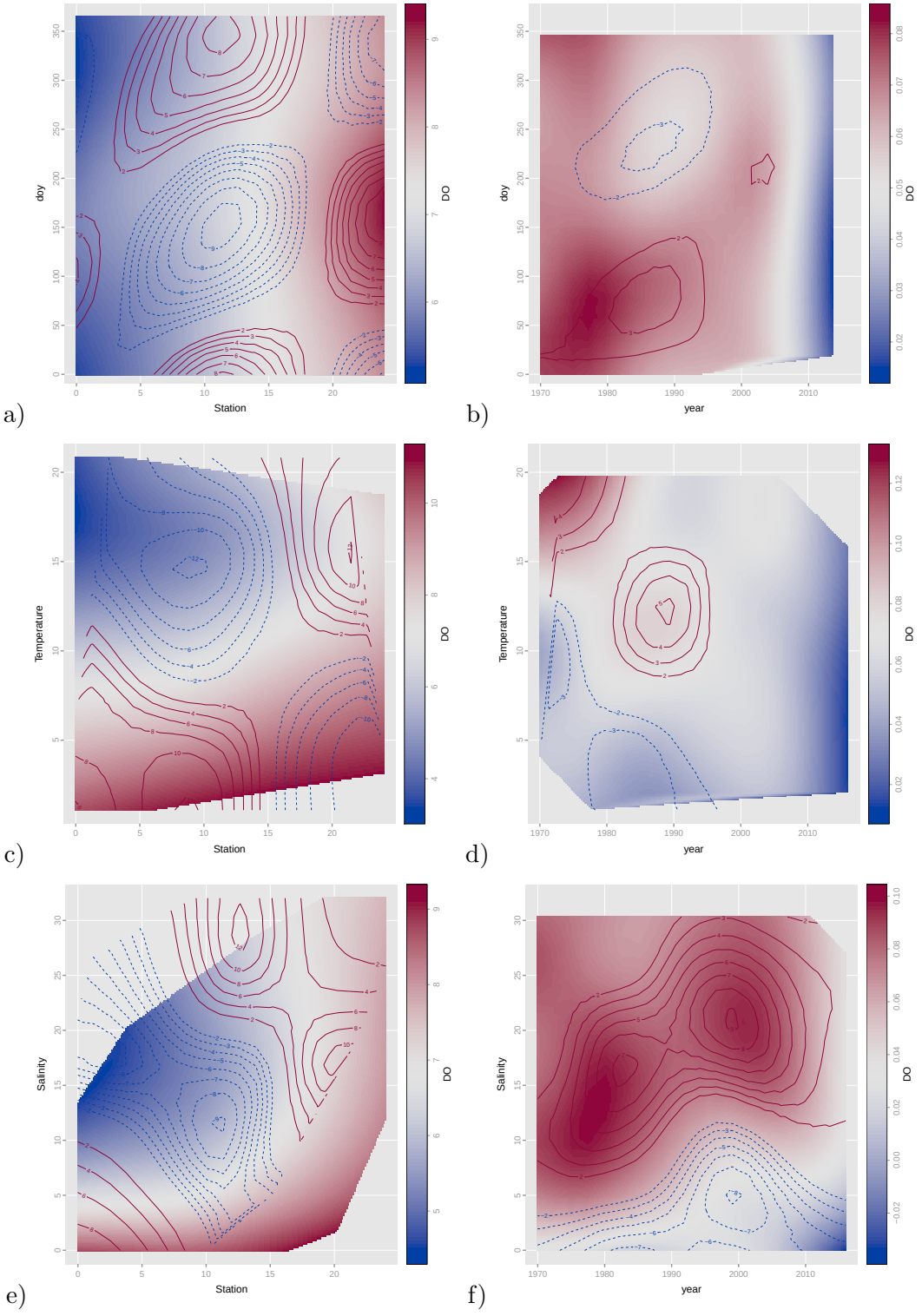


FIGURE 7.2: Interaction plots of a) Day of Year : Station, b) Day of Year : Year, c) Temperature : Station, d) Temperature : Year, e) Salinity : Station, and f) Salinity : Year for derivative model.

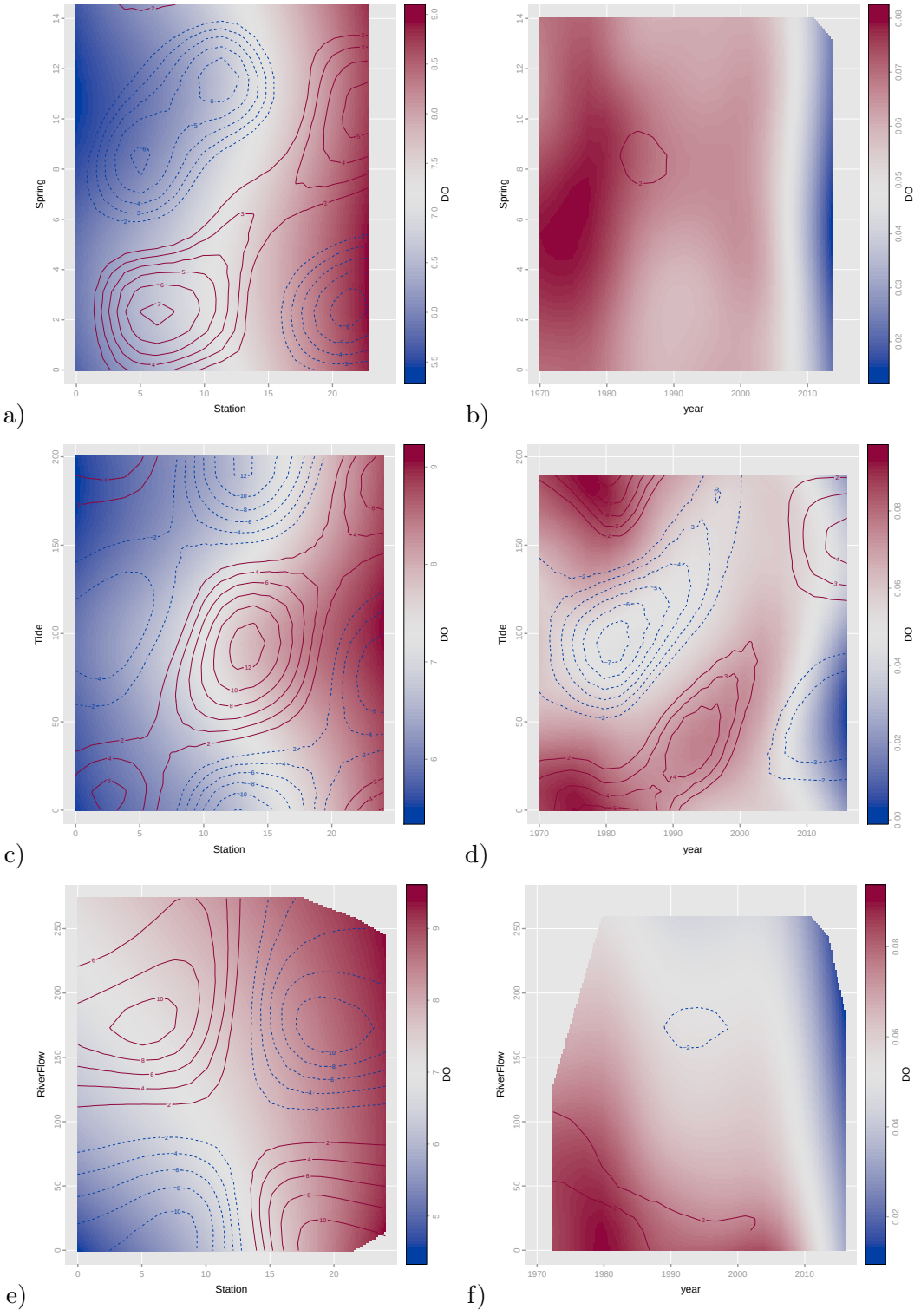


FIGURE 7.3: Interaction plots of a) Spring : Station, b) Spring : Year, c) Tide : Station, d) Tide : Year, e) River Flow : Station, and f) River Flow : Year for derivative model.

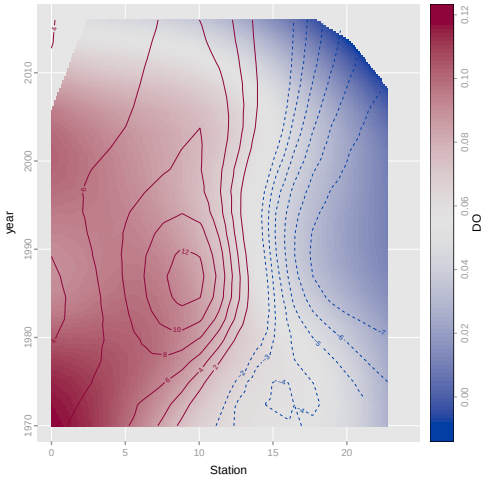


FIGURE 7.4: Interaction plot of Year : Station for derivative model.

7.1.2 Trivariate Interaction Terms

The trivariate interaction terms of the derivative model offer an acute sensitivity for detecting subtle yet sudden increases in the dissolved oxygen. Contour plots depict the overall partial derivative with respect to Year via the underlying colours and the adjustments made to the partial derivative with respect to Year via the contours. Figures 7.5-7.10 depict all 6 trivariate interactions with Year values of 1970, 1973, 1977, 1980, 1983, 1987, 1990, 1993, 1997, 2000, 2003, 2007, 2010, 2013, and 2016. Together with Station along the vertical axis, one can see where and when the dissolved oxygen was increasing at more rapid rates. The various covariates along the horizontal axis allow one to see the effect these covariates have on the rate of increase for a specific time and place. Some of these plots are more informative than others. For example, the Day of Year : Station : Year interaction for the derivative model not only pinpoints the Year and Station values of more rapid rates of increased dissolved oxygen, but also what time of the year these rapid rates of increase are occurring. Figure 7.5 depicts a rapid rate of increase of dissolved oxygen at Stations 0 through 10 in the 1970's and 1980's that starts to dissipate in the 1990's and 2000's. There is also a significant positive adjustment at Station 12 which starts to develop in late Spring 1983 and drifts down to Station 14 by 2003.

There is still value in knowing how the other covariates affect the rate of increase at a particular time and place. There is a wealth of information contained in these figures, much of which requires comprehensive understanding of river systems, specifically the River Clyde, would be able to draw more insightful conclusions than the ones presented here. For this reason, the code and detailed guidelines to regenerate this derivative model and accossiated plots has been given to SEPA.

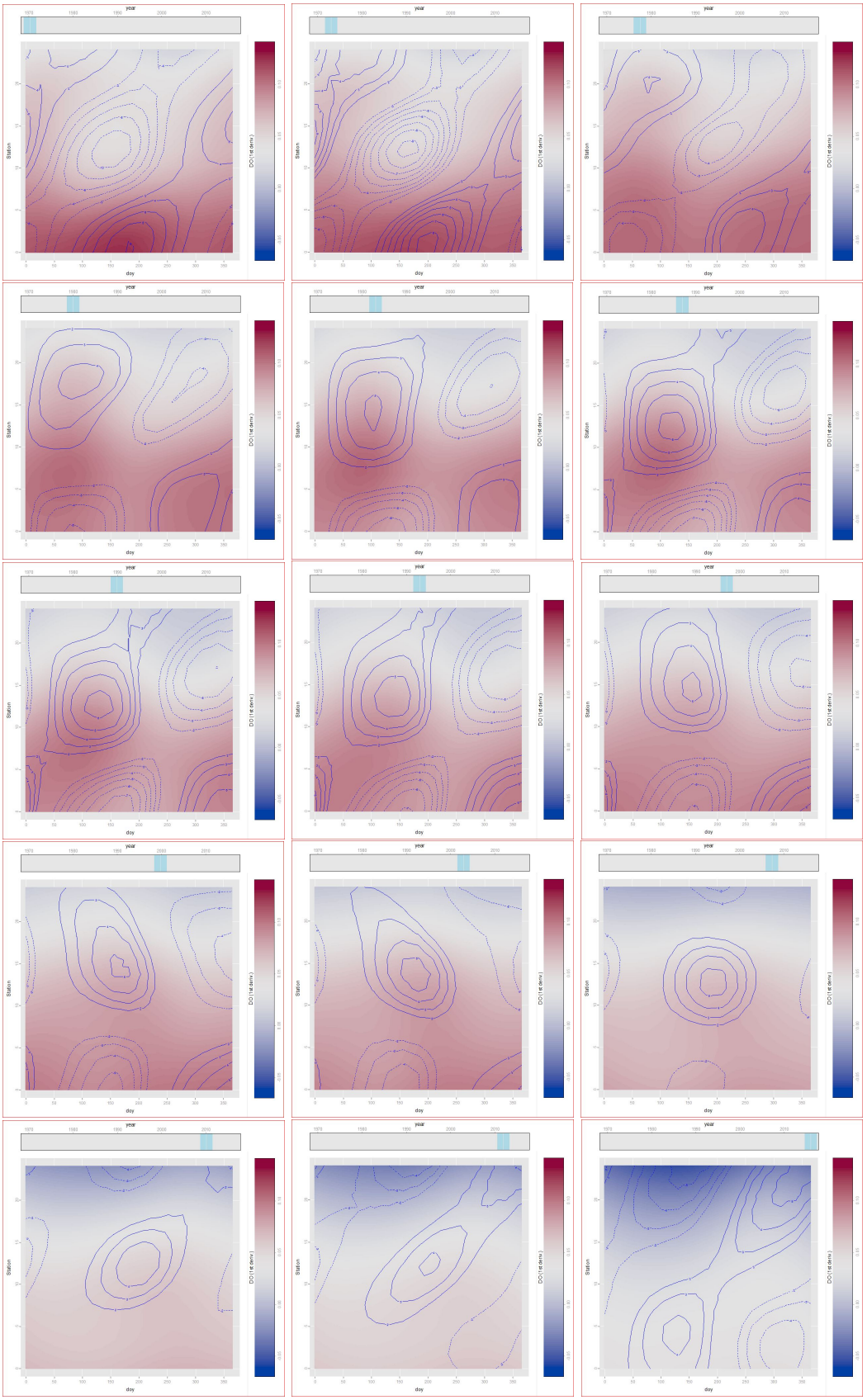


FIGURE 7.5: Plots of Day of Year : Station : Year interaction of derivative model with Year varied across top axis.

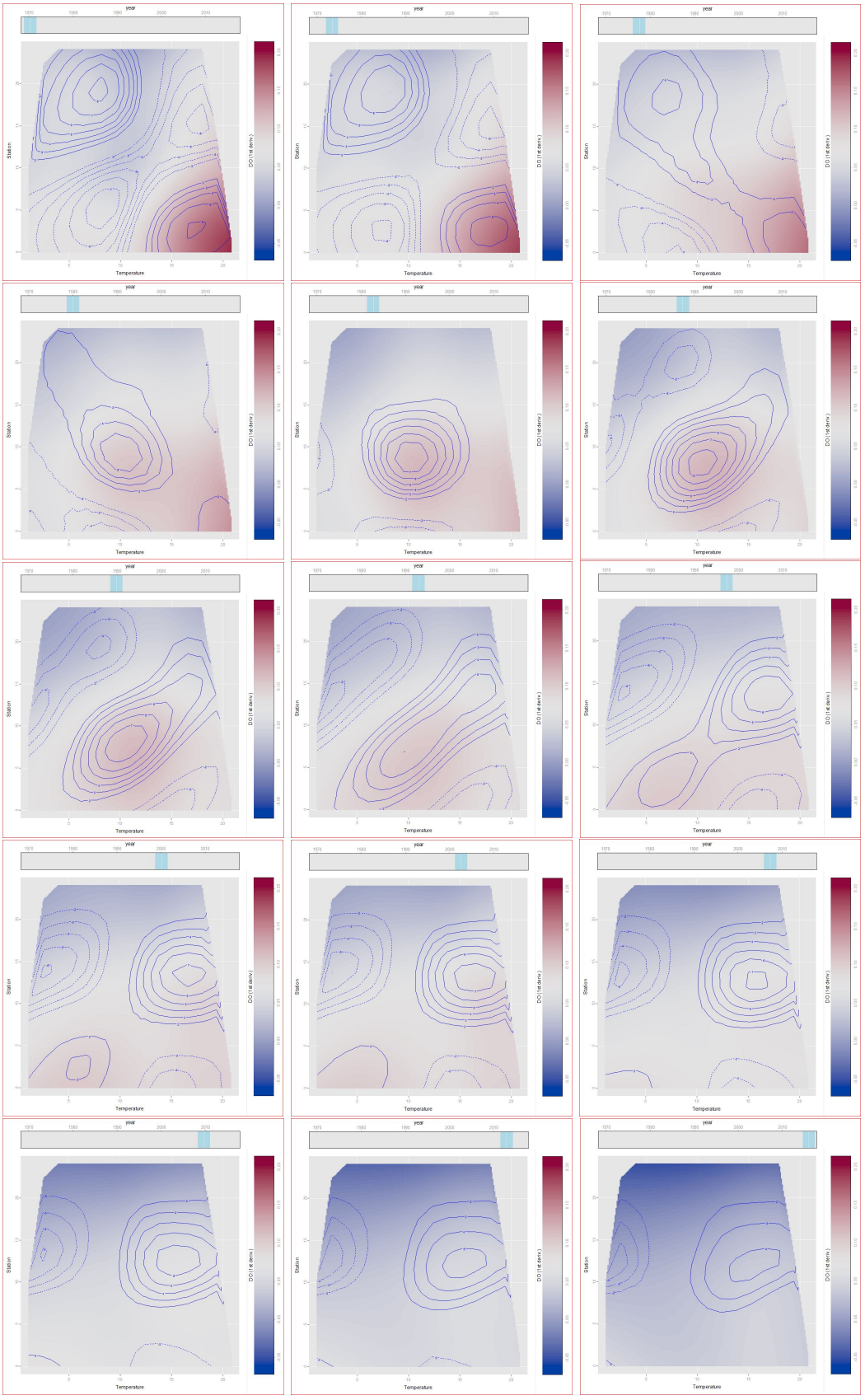


FIGURE 7.6: Plots of Temperature : Station : Year interaction of derivative model with Year varied across top axis.

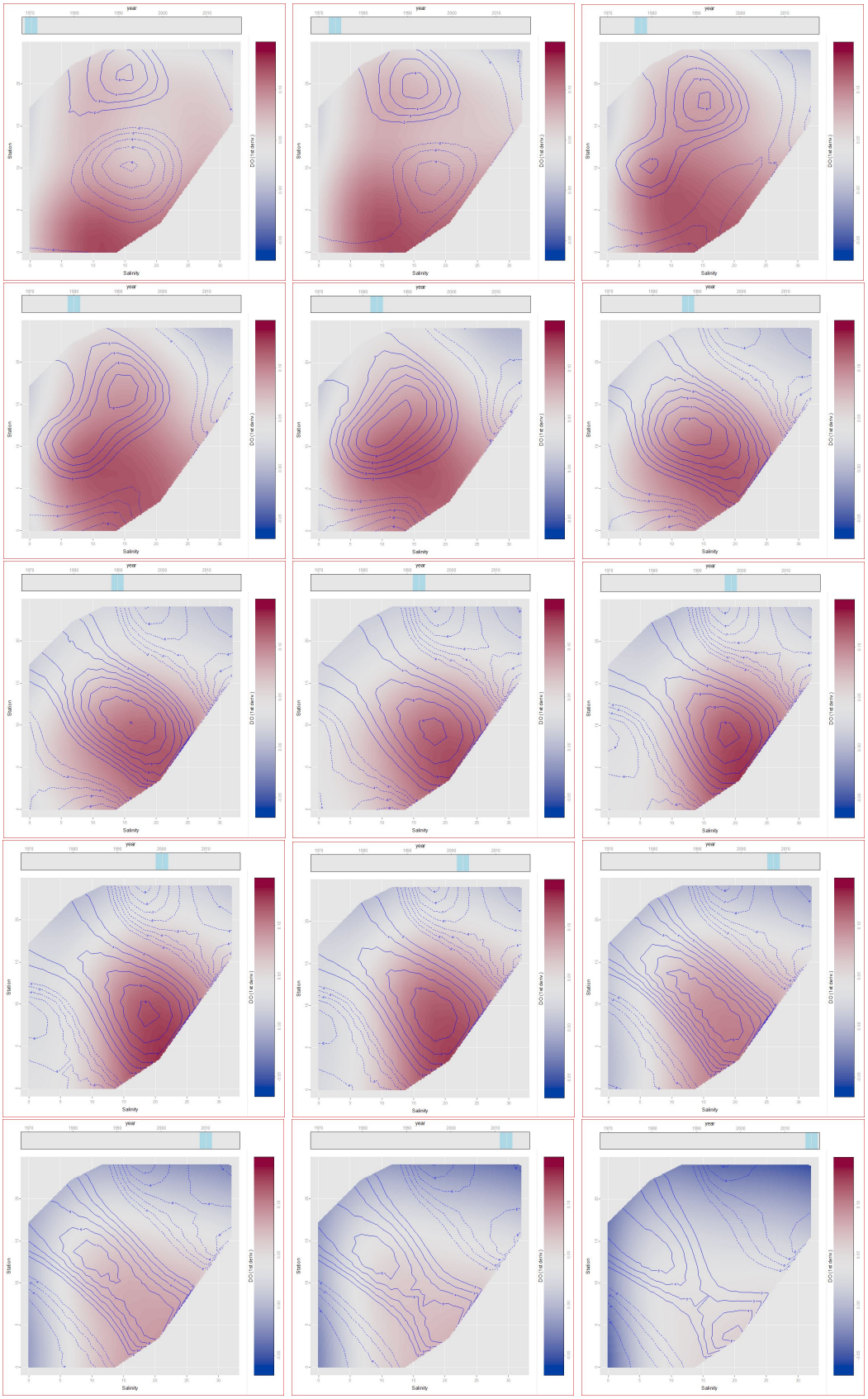


FIGURE 7.7: Plots of Salinity : Station : Year interaction of derivative model with Year varied across top axis.

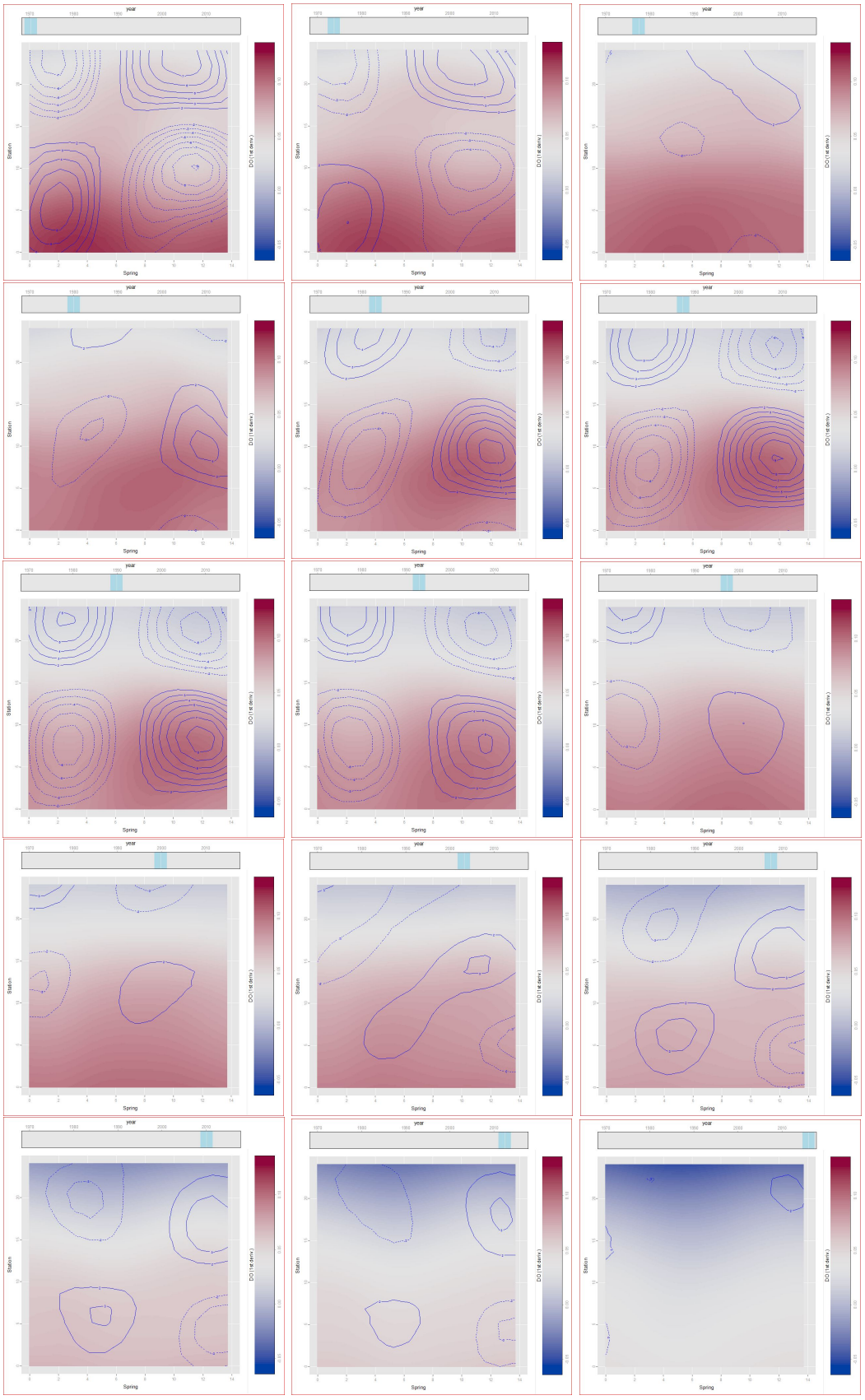


FIGURE 7.8: Plots of Spring : Station : Year interaction of derivative model with Year varied across top axis.

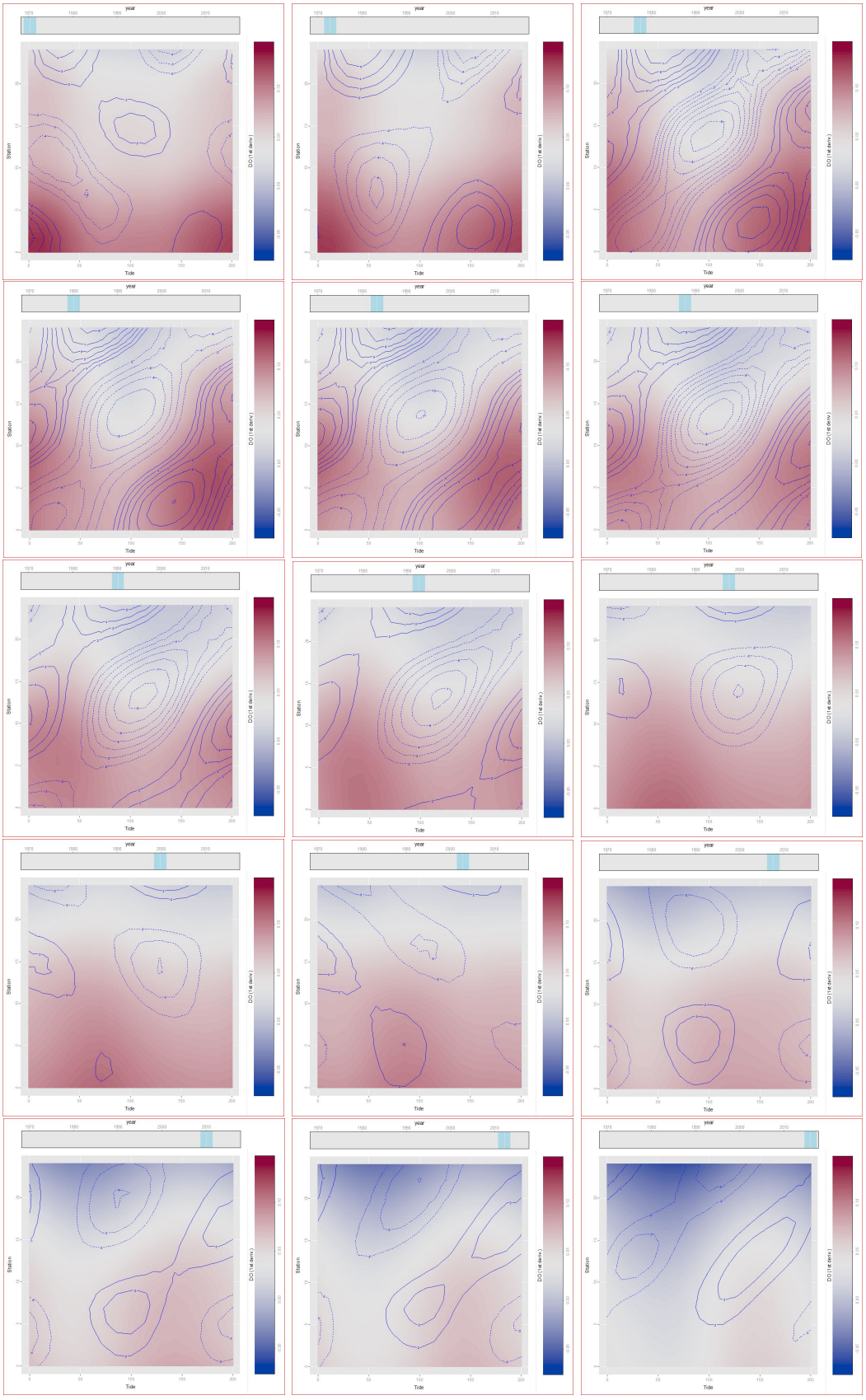


FIGURE 7.9: Plots of Tide : Station : Year interaction of derivative model with Year varied across top axis.

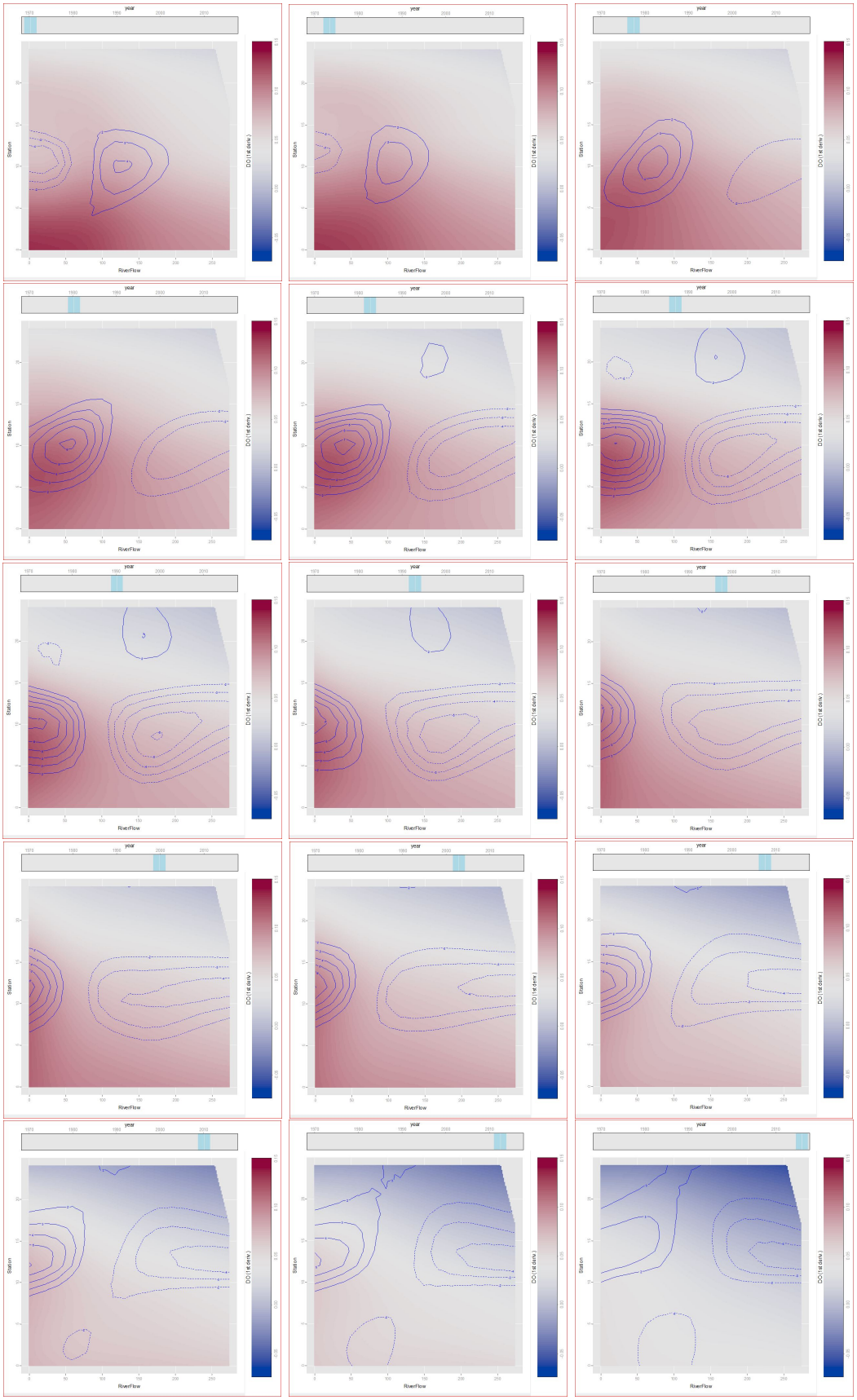


FIGURE 7.10: Plots of River Flow : Station : Year interaction of derivative model with Year varied across top axis.

7.2 Impacts of Shielhall and Dalmuir Wastewater Treatment Facility Upgrades

Having a representative statistical model for a system as complex as the River Clyde not only allows for the identification of drivers of the dissolved oxygen levels, but also facilitates the detection of positive impacts brought on by known events in a steadily improving system. In particular, the derivative model takes into account many potential influences, as well as interactions of those potential influences with Station and Year, thus allowing for the possible detection of the 1985 Shieldhall upgrade and 2001-2004 Dalmuir upgrade.

Figure 7.11 is comprised of two perspective plots and a contour plot of the Day of Year : Station : Year interaction of the derivative model for Year 1985, the year of the Shieldhall upgrade. The two perspective plots have been rotated about the DO (dissolved oxygen) axis so that there is a clear view of the surface with a line of sight perpendicular to the Day of Year (doy) axis in the top panel and a line of site perpendicular to the Station axis in the middle panel. A local maximum of the partial derivative occurs at approximately Day of Year 100 between Station 6 and 8 with significant positive adjustments at Station 12. The Shieldhall wastewater treatment facility is located at Station 4, just 2 to 4 miles upstream of the local maximum. This is compelling evidence the upgrade had a significant positive effect on the dissolved oxygen.

Figure 7.12 is set up the same way as Figure 7.11 with Year set to 2003, within the time interval of the Dalmuir upgrade (2001 to 2004). A local maximum of the partial derivative occurs at approximately Day of Year 200 at Station 10 with significant positive adjustments at Station 14. Dalmuir wastewater treatment facility is located at Station 8, just 2 miles upstream of the local maximum. Here too is compelling evidence the upgrade had a significant positive effect on the dissolved oxygen. The effect was not as abrupt as it was for Shieldhall, but this may be due to the slower implementation of the upgrade.

Figure 7.13 offers a different take on the realization of the positive effects brought on by the upgrades. The variable across the top axis is Station with Year along the horizontal axis and Day of Year (doy) along the vertical axis. Paying specific attention to the 3^{rd} row of panels, it is evident significant positive adjustments are developing between Stations 8 and 14 during Year 1985 and 2005. These results are consistent with the evidence shown above.

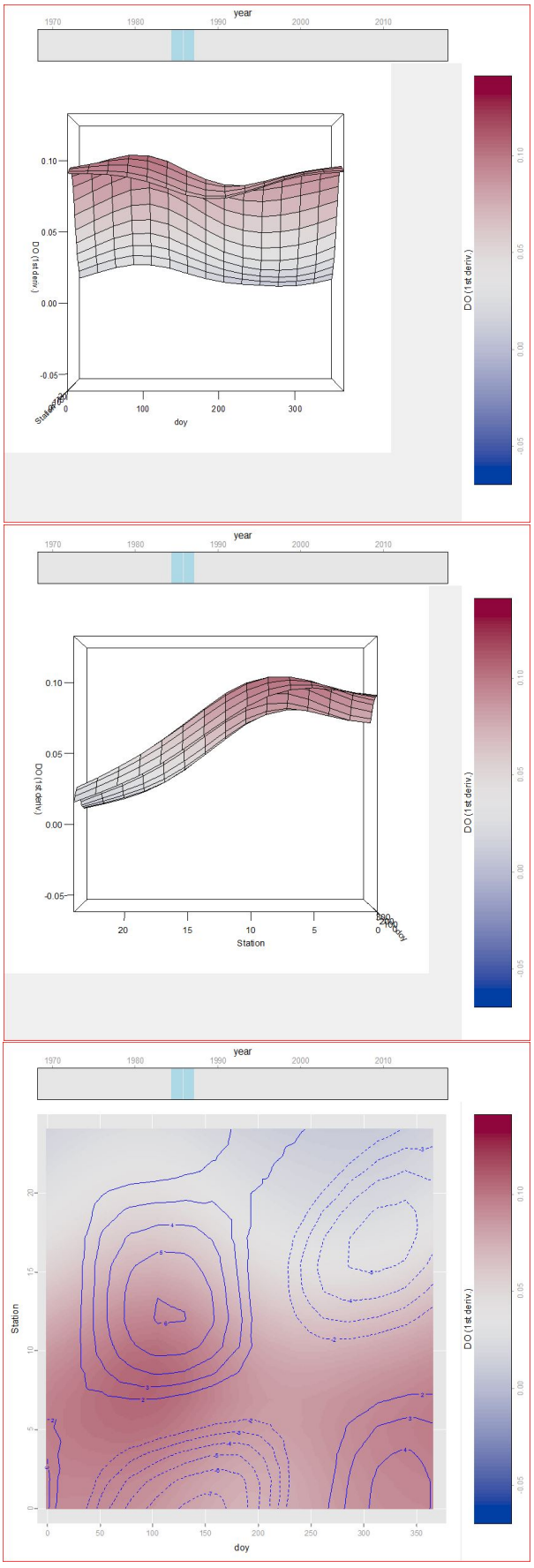


FIGURE 7.11: Plots of Day of Year : Station : Year interaction of derivative model with Year=1985 representing the upgrade at Shieldhall.

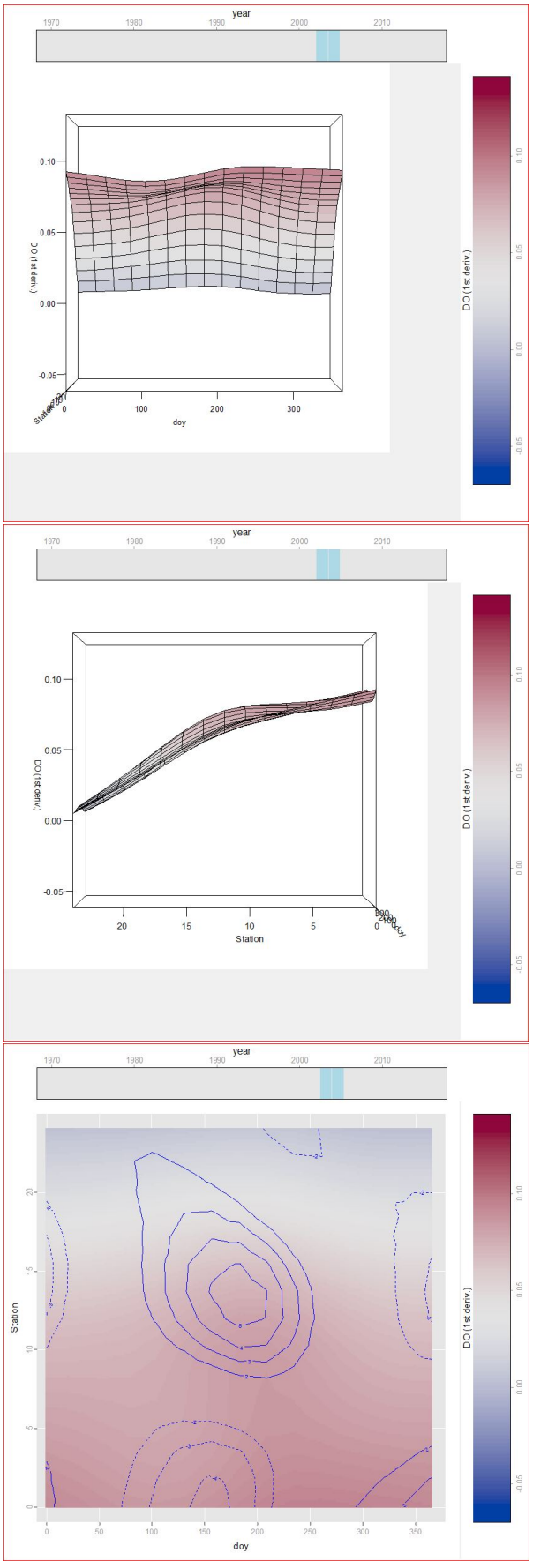


FIGURE 7.12: Plots of Day of Year : Station : Year interaction of derivative model with Year=2003 representing the upgrade at Dalmuir.

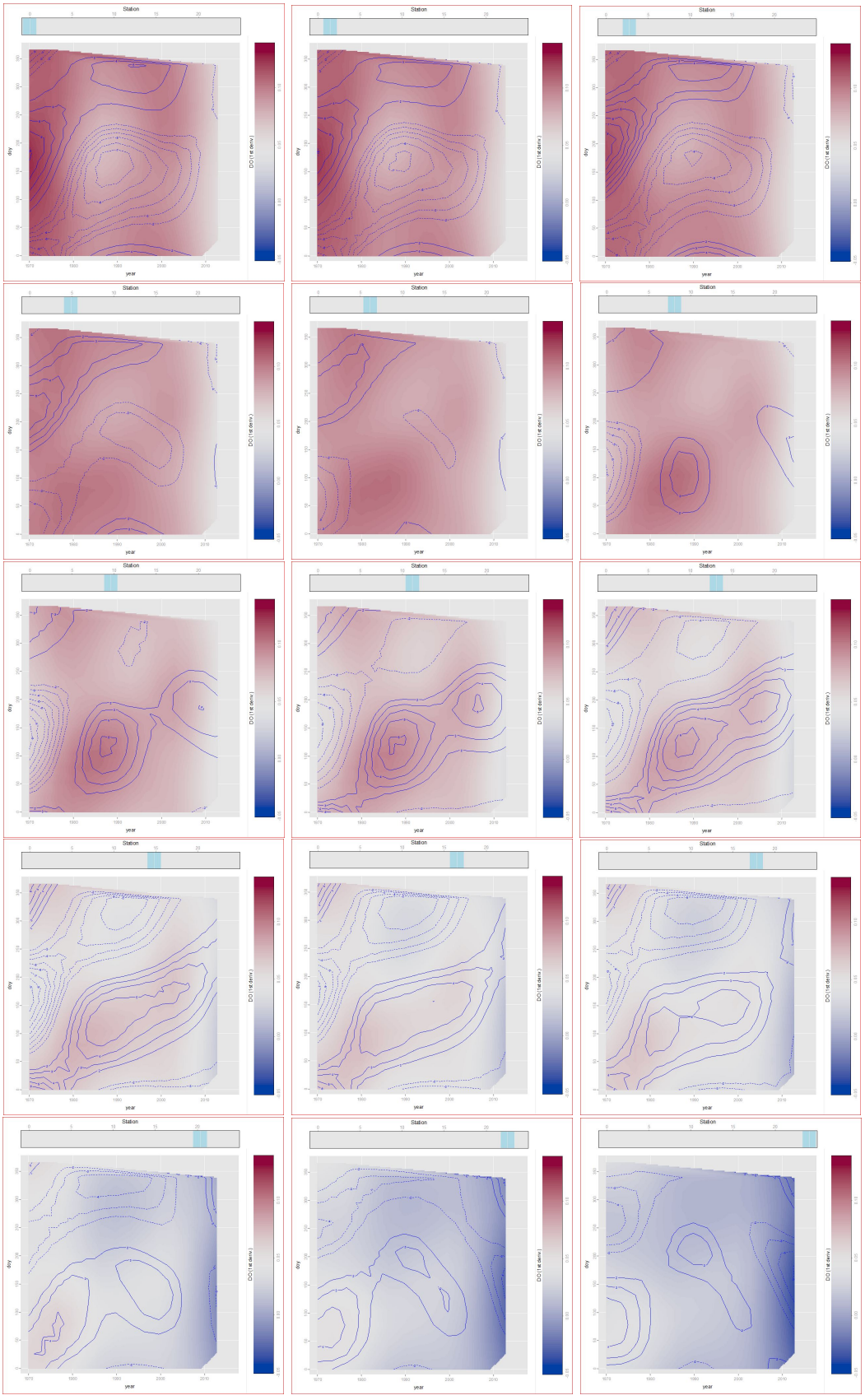


FIGURE 7.13: Plots of Day of Year : Station : Year interaction of derivative model with Station varied across top axis.

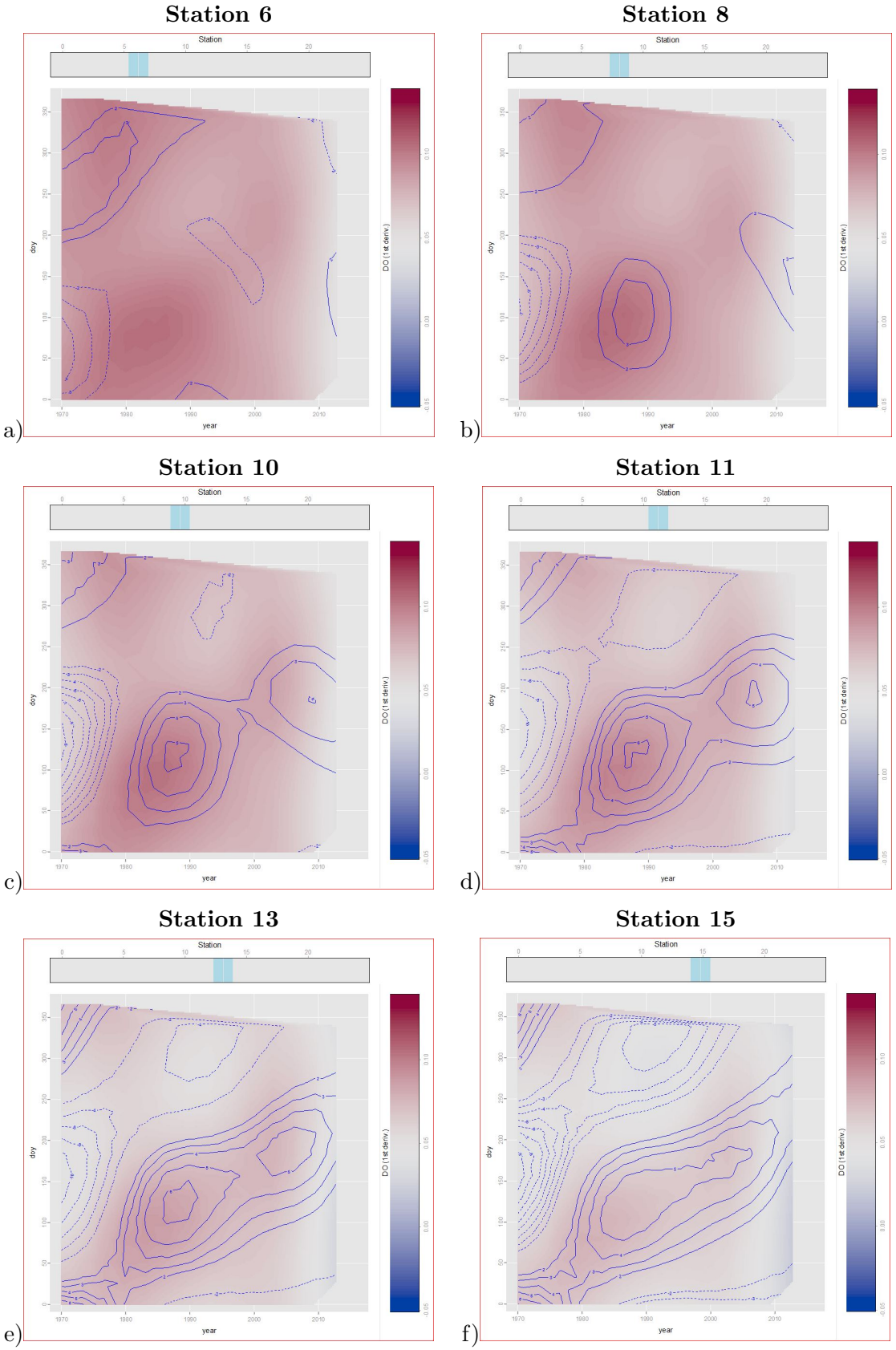


FIGURE 7.14: Plots of Day of Year : Station : Year interaction of derivative model for Stations a) 6, b) 8, c) 10, d) 11, e) 13, and f) 15.

A closer look at the positive changes which may be attributed to the Dalmuir and Shieldhall upgrades can be seen in Figure 7.14. It is important to note the contours of these plots are depicting the adjustment the Day of Year : Station : Year interaction is making to the underlying main effects and bivariate interaction terms for the partial derivative with respect to Year, whereas the background colors are depicting the partial derivative itself. Although the actual increase in dissolved oxygen cannot be determined by these plots, the plots do show where and when the rate of increase was large. Positive adjustment contours start to appear at Station 8 with background colour depicting the highest partial derivative at Day of Year 75 and Year at 1985. Station 10 and 11 are seeing stronger positive adjustment contours developing circa 1986 and the start of positive adjustment contours developing circa 2006. By Station 13, there are clear positive adjustment contours circa 1985 and 2005. These contours start to dissipate by Station 15. The bottom right panel of Figure 7.13 shows these positive adjustment contours have almost disappeared by Station 24. Unlike with the contours which depict the adjustment, the background color depicts the derivative itself. A darker maroon color corresponding to Day of Year 75 and Year 1985, the year of the Shieldhall upgrade, is clearly present at Stations 6, 8, 10, and 11. This suggests there was an increase in the positive rate of change of the dissolved oxygen. The same is happening, to a lesser degree, at Day of Year 2000 and Year 2006, the year of the Dalmuir upgrade, for Stations 11 and 13. These results are also consistent with the evidence shown above.

7.3 Summary

The derivative additive mixed model with interactions fitted in Chapter 7 differs from the additive mixed model with interactions fitted in Chapter 5 in that the penalty order for Year was changed from 2 to 3. Chapter 6 results showed that penalty order 3 outperformed other penalty orders when the main objective was derivative estimation. The derivative additive mixed model with interactions had smoothing parameters selected by BIC in the same fashion as the additive mixed model with interaction. Plots of the main effects and interaction terms of the derivative model were presented and interpreted, many of which did not differ strongly from the non-derivative model. The main differences were seen when the explanatory covariate of Year was involved and the first derivative with respect to Year and the first partial derivatives with respect to year were taken.

Special attention was paid to the partial derivative of the Day of Year : Station : Year interaction. This interaction was the one that would allow for the detection of any subtle increase in the rate of change of dissolved oxygen for a particular Year and

Station. Although the Shielhall and Dalmuir wastewater treatment facilities are located at Station 4 and 8 respectively, it is reasonable to assume the impacts, if any, of the upgrades would be present a short distance down stream. The derivative additive mixed model with interactions reinforces the evidence that the upgrades had a positive impact.

Chapter 8

Conclusion

One novelty presented in this thesis is the simulation study conducted in Chapter 4 which analyzed the calibration of the analysis of variance methods contained in the `sm` package. The simple F-test and the quadratic forms test are two options for ANOVA. These methods were compared to a computationally expensive parametric bootstrap method. The simulation study showed the quadratic forms method of additive model selection performed the best by showing superior size compared to the simple F-test and superior power when compared to the parametric bootstrap. Thus, the quadratic forms method is well calibrated and has low computational cost. Also, the `sm` methods outperformed the `mgcv` methods by displaying superior power.

P-spline additive mixed models with interactions are by no means novel. What is novel in this thesis is the complexity of the River Run data and additive mixed models with interactions fitted to that data in Chapter 5. There are 8 explanatory covariates, each represented by a main effect. Year and Station are each combined with the remaining 6 covariates and each other for a total of 13 bivariate interaction terms. Furthermore, Year and Station together are combined with the remaining 6 covariates for a total of 6 trivariate interaction terms. The sampling regime where multiple samples are taken during the same day justifies the inclusion of a random effect. This results in an additive mixed model comprised of 27 fixed effect terms and a random effect term. Adding to the complexity was the task of selecting smoothing parameters for each of the main effects. Chapter 5 showed BIC score outperformed GCV, AIC, and AICc for smoothing parameter selection for additive models and additive mixed models for the River Run data. The other scores had difficulties achieving minimum values. The sequence of degrees of freedom optimization of the main effects was in question. A simulation study involving 5 bivariate additive models with interactions was conducted with results showing the said order had little effect on the final degrees of freedom of each main effect. This result may

not be necessarily extended to more complex additive mixed models with interactions. The justification of tuning the Year and Station smoothing parameters last was that one, the other, or both of these covariates were included in every interaction. It seems plausible tuning the Year and Station smoothing parameters would influence every other main effect. Furthermore, 3 other permutations were used in optimizing the degrees of freedom and liitle change was seen from the original permutation. The inclusion of interactions tended to reduce the optimal degrees of freedom. The inclusion of random effects also tended to reduce some degrees of freedom. The resulting complex mixed model helps with the interpretation of the interaction via interaction plots.

The research presented in this thesis has produced numerous positive outcomes. Arguably, the most impressive novel component of this thesis was presented in Chapter 6. Penalty order selection does make a difference when derivative estimation is the primary goal. The proposal, justification, and subsequent verification that P-spline penalty order 3 outperforms the customary penalty order 2 when estimating derivatives. A wide variety of univariate and bivariate functions were considered and p3 was the optimal penalty order for the univariate case. The bivariate case found p23 to be optimal for 500 data points and p3 to be optimal for 5000 data points. Since the River Run data had more than 7000 points, p3 was subsequently used for the derivative model in Chapter 7.

Chapter 7 detailed the construction of an additive mixed model with interactions as was done in Chapter 5, with the difference of changing the Year term's penalty order from 2 to 3. This was done in light of the results in Chapter 6 for the purpose of generating a derivative with respect to Year model. Chapter 7 showed partial derivatives are effective in detecting sudden changes in a system which is steadily improving. It was shown that the upgrades to the Shieldhall and Dalmuir wastewater treatment facilities had detectable positive impacts about 4 miles down stream of those facilities. This was evident in Figures 7.11, 7.12, and 7.14

There is potential to expand on the work presented in this thesis. Some suggestions for future work are as follows:

- Address the issue of model selection for additive mixed models
- Explore other methods of smoothing parameter selection.
- The modeling performed here can be applied to different depths.
- Explore the impact of other known events, possibly detrimental, to assess the detection performance of the derivative model
- Apply these methods to other rivers where similar data has been collected, or other systems not necessarily dealing with rivers.

- A further problem arises in the analysis of Inner Clyde Estuary (ICE) buoy data and possibly fitting a additive model with an auto-regressive correlation component. There are also other fields in the buoy data, such as pH and turbidity not present in the River Run. There exists an opportunity for a more complex model.

Here is a little background on the ICE buoy data set. SEPA has recently deployed monitoring and data collecting buoys at several locations along the River Clyde. One specific buoy deployed in 2011 is the Inner Clyde Estuary (ICE) Buoy. Figure 8.1 shows the buoy's location just west of the Riverside Museum. This location roughly corresponds to mile 2 of the River Run. The ICE buoy collects data from this single location of the river. Data are recorded at 15 minute intervals, although there are several gaps in data collection due to technical difficulties. Data are collected from the surface and the bottom of the river.

The ICE buoy data spans a much smaller time interval than the river run data. The buoy data also lack the capability of generating a model which includes a spatial component such as is included in the River Run model. It is however dense with respect to time. All fields used in the River Run data are represented in some form in the ICE buoy data.

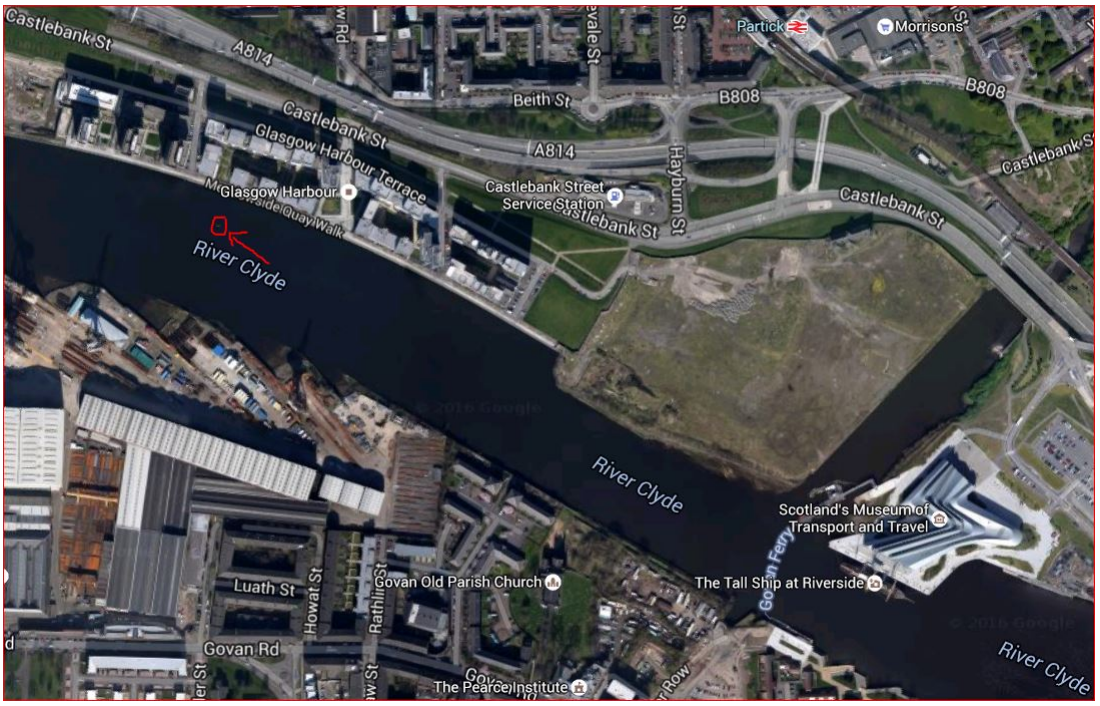


FIGURE 8.1: The map above depicts the position of the ICE buoy

There are also some added fields which create an opportunity for more complex models to be generated in the future. Once again while most of the fields in the ICE buoy data are self explanatory, some fields warrant further description. These are

Data Field	Units of Measurement (Factor Assignment *)
Dissolved Oxygen	milligrams per liter
Temperature	degrees Celsius
Salinity	parts per thousand
Date of Sample	time and date
Tide State (Lunar Semi-Dihernal or M2)	ebb or flood *
Tide Fraction (M2)	percentage
Springiness	pure number (no units)
River Flow	cubic meters per second
Chlorophyll	micrograms per liter
Turbidity	Nephelometric Turbidity Unit
pH	pH units
Tide Height	meters
Rainfall	millimeters

TABLE 8.1: The table above shows the data fields collected by ICE buoy.

- Tide Fraction (M2) - a value from a linear scale on the interval [0,100], where 0 corresponds to max/min tide and 100 corresponds to the subsequent max/min tide.
- Springiness - a value from a linear scale on the interval [0,1], where 0 corresponds to neap tide and 1 corresponds to spring tide.
- River Flow - different from the River Run, these values are actual readings taken every 15 minutes.

As is true for the River Run data, there is a need to create new data fields from the raw data for the purpose of extracting as much information as possible and to transform certain raw data fields into fields conducive to statistical model construction.

- Day of Year - a whole number value from the cyclic interval [1,365]
- Year - a value with a fractional component to constitute a continuous time scale
- Tide (M2) - a value from a linear scale on the interval [0,200], where 0 and 200 corresponds to low tide and 100 corresponds to high tide.
- Spring - a value from a linear scale on the interval [-1,1], where 0 corresponds to neap tide and -1 and 1 corresponds to spring tide.

As the two data sets have many data fields in common, there is a valuable opportunity to compare models from each data set. The models cannot have exactly the same structure due to the inherent differences in sampling. The River Run model required a random effects component corresponding to sample date as measurement at all sampling stations are made on the same day. The ICE Buoy would require an auto-regressive correlation

component due to the 15 minute time gap between sampling. The other striking difference between data sets is that the River Run has multiple sampling locations whereas the ICE buoy has a single sampling point. Nevertheless, restricting the River Run model predictions to the appropriate sampling stations would allow one to compare the main effects and interaction terms of the River Run model to those of the ICE buoy model.

Appendix A

Plots of Derivative Estimation Deviation

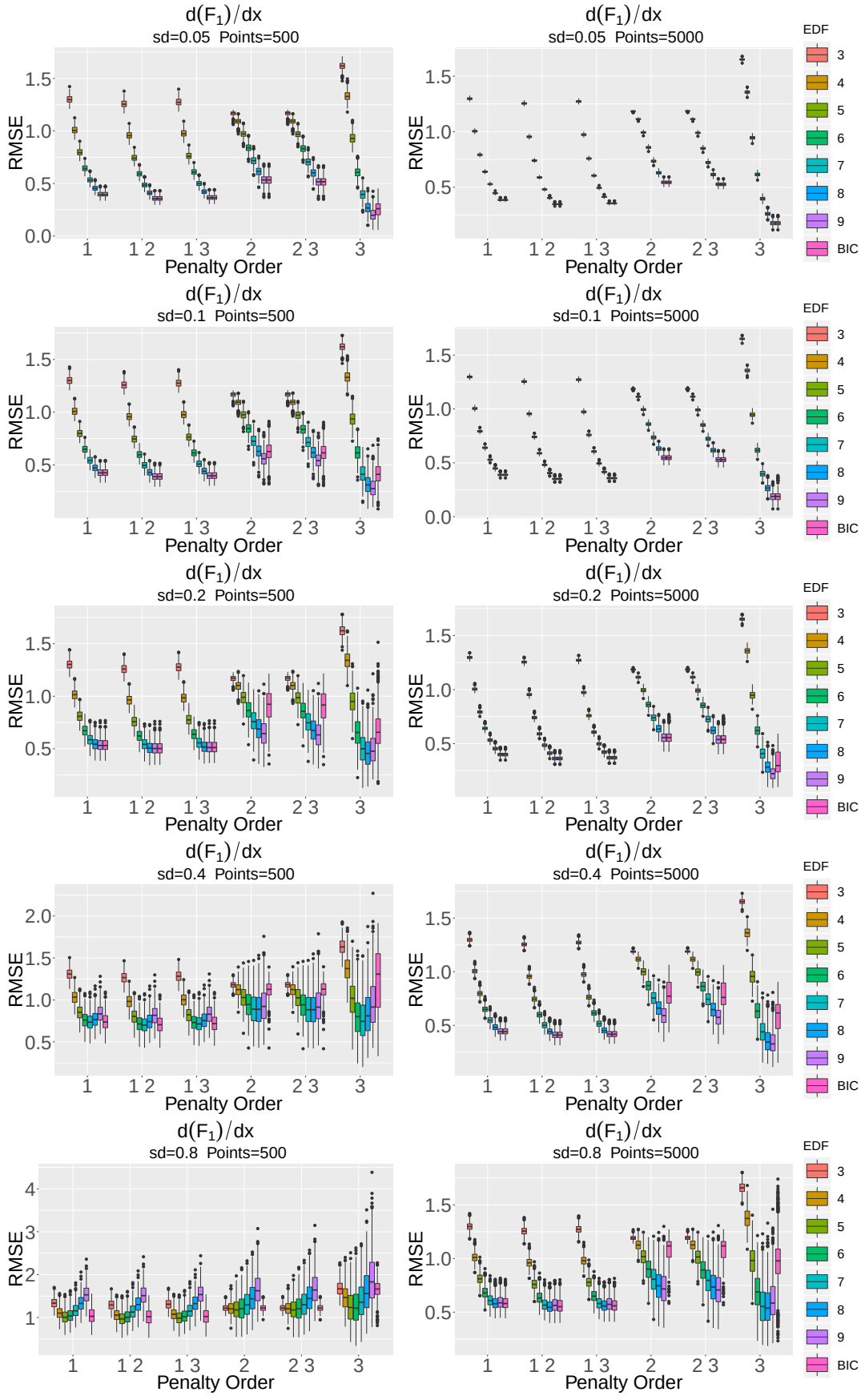


FIGURE A.1: Box plots of $\frac{d(F_1)}{dx}$ estimation RMSE for 500 and 5000 data points.

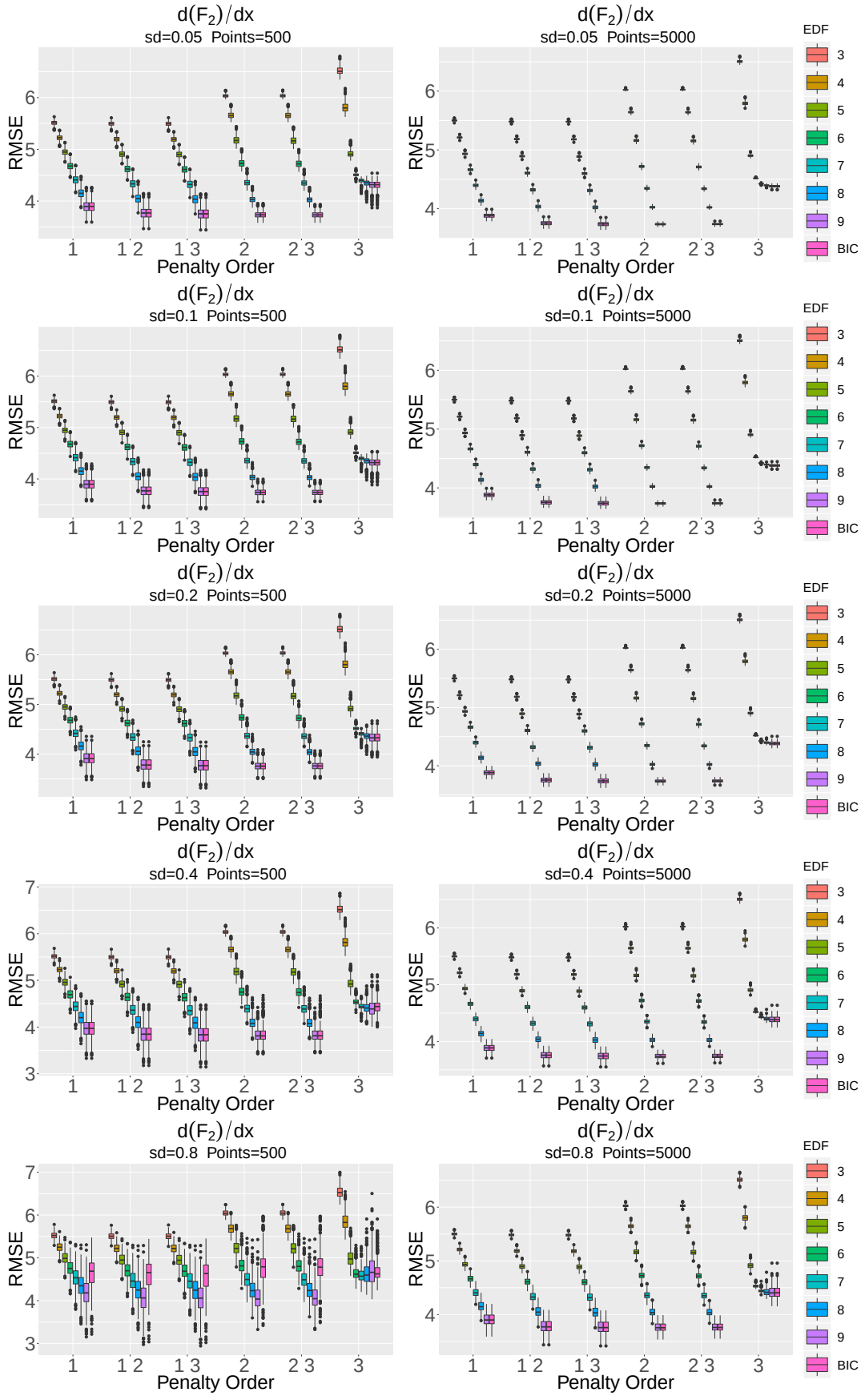


FIGURE A.2: Box plots of $\frac{d(F_2)}{dx}$ estimation RMSE for 500 and 5000 data points.

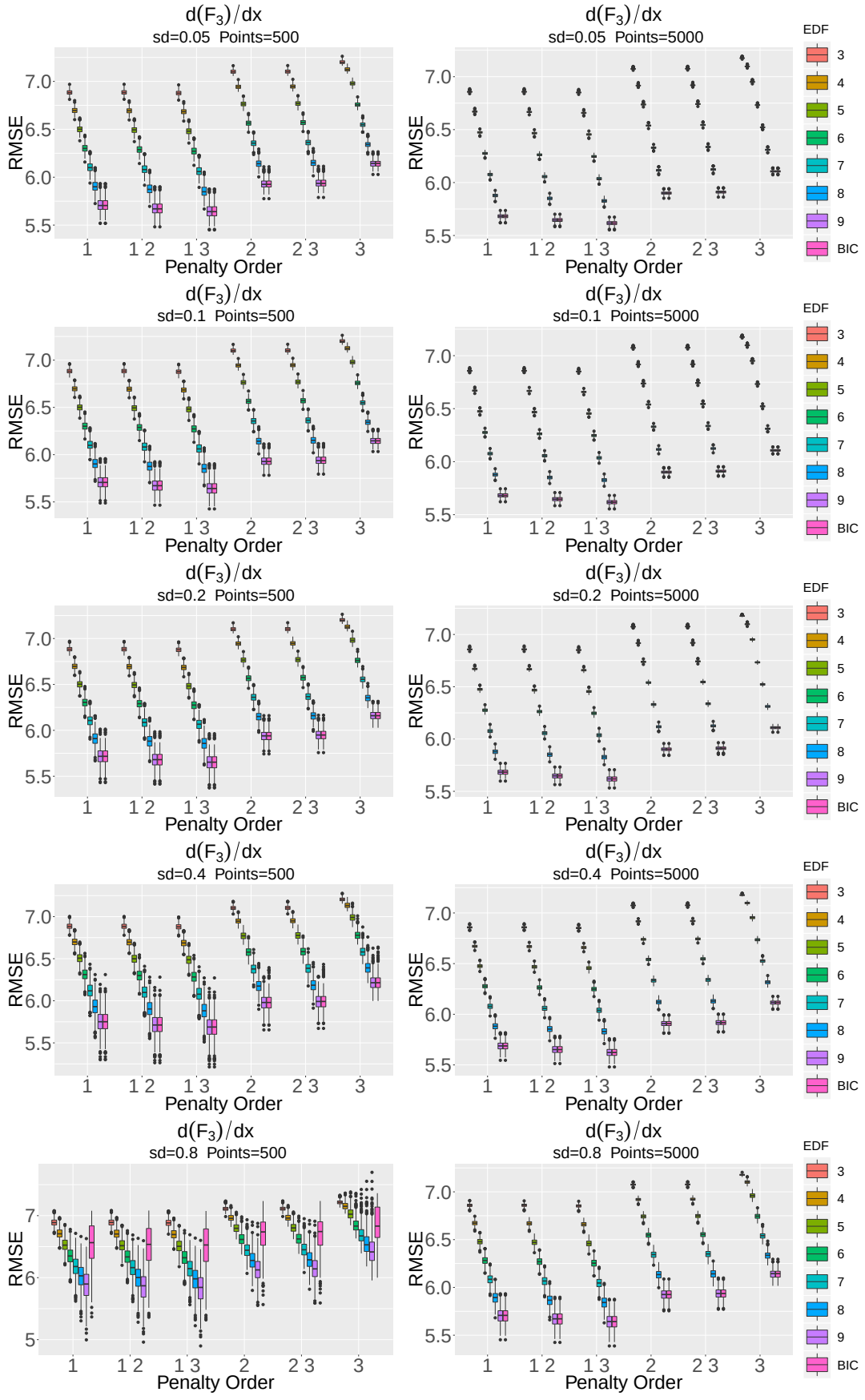


FIGURE A.3: Box plots of $\frac{d(F_3)}{dx}$ estimation RMSE for 500 and 5000 data points.

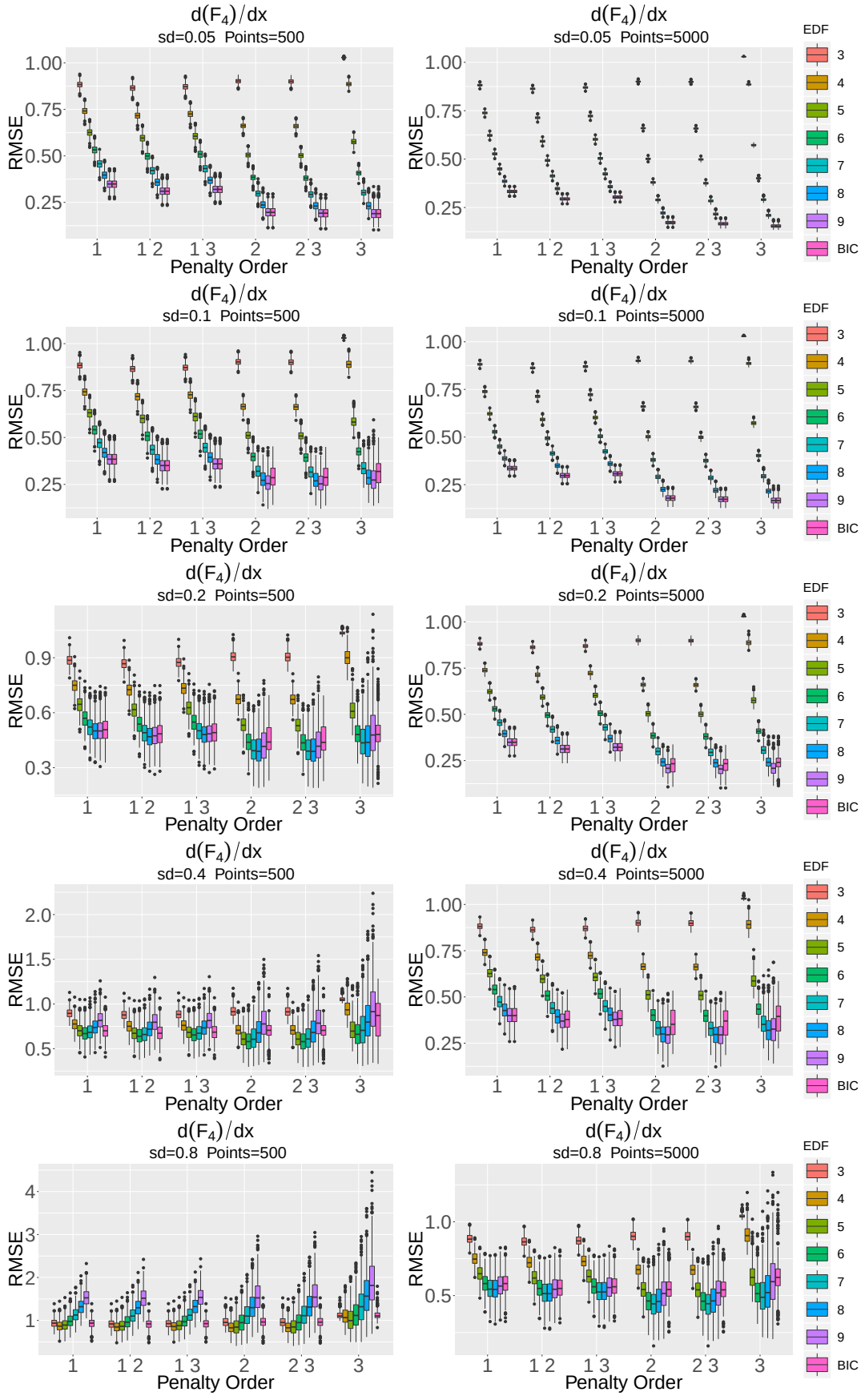


FIGURE A.4: Box plots of $\frac{d(F_4)}{dx}$ estimation RMSE for 500 and 5000 data points.

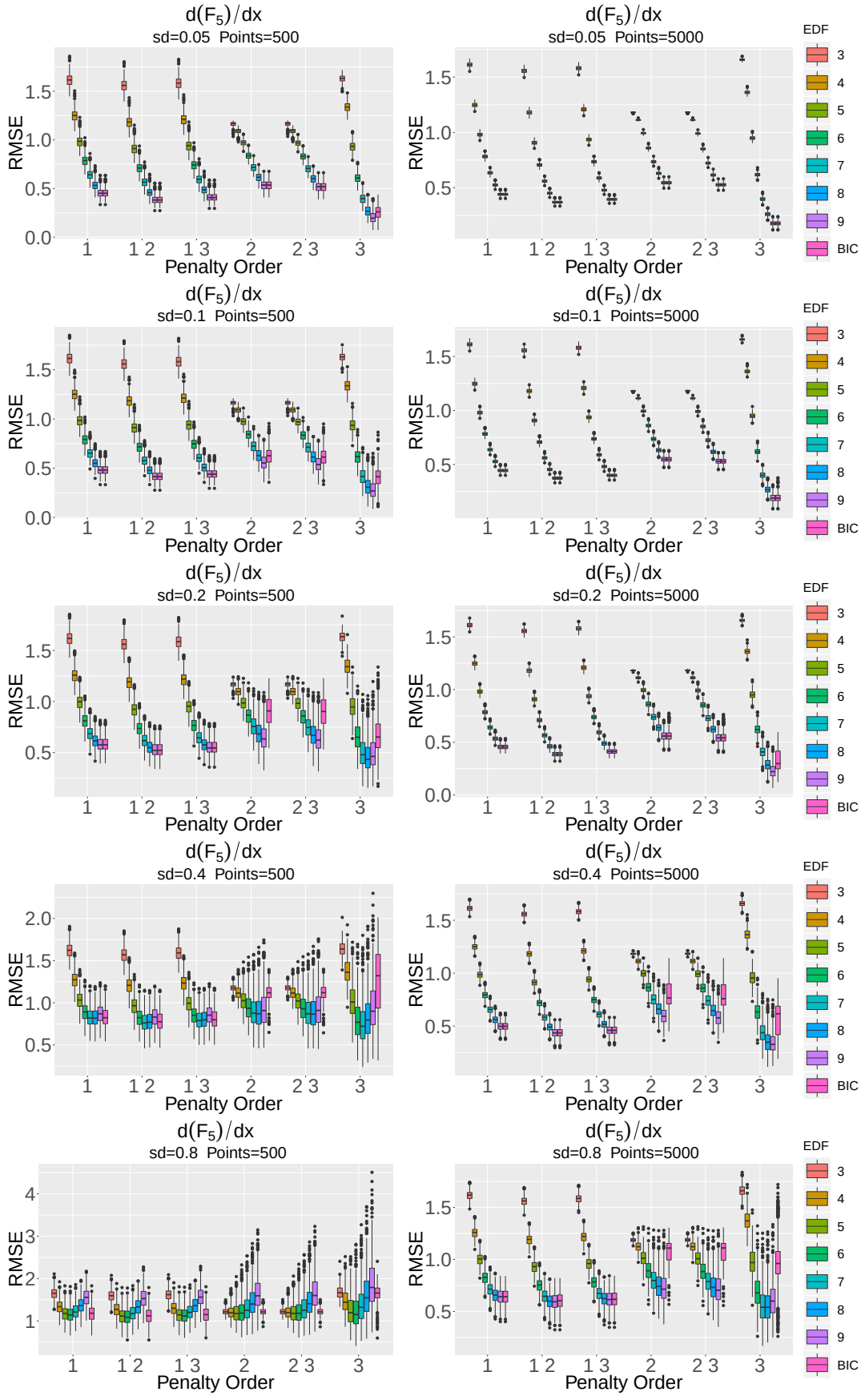


FIGURE A.5: Box plots of $\frac{d(F_5)}{dx}$ estimation RMSE for 500 and 5000 data points.

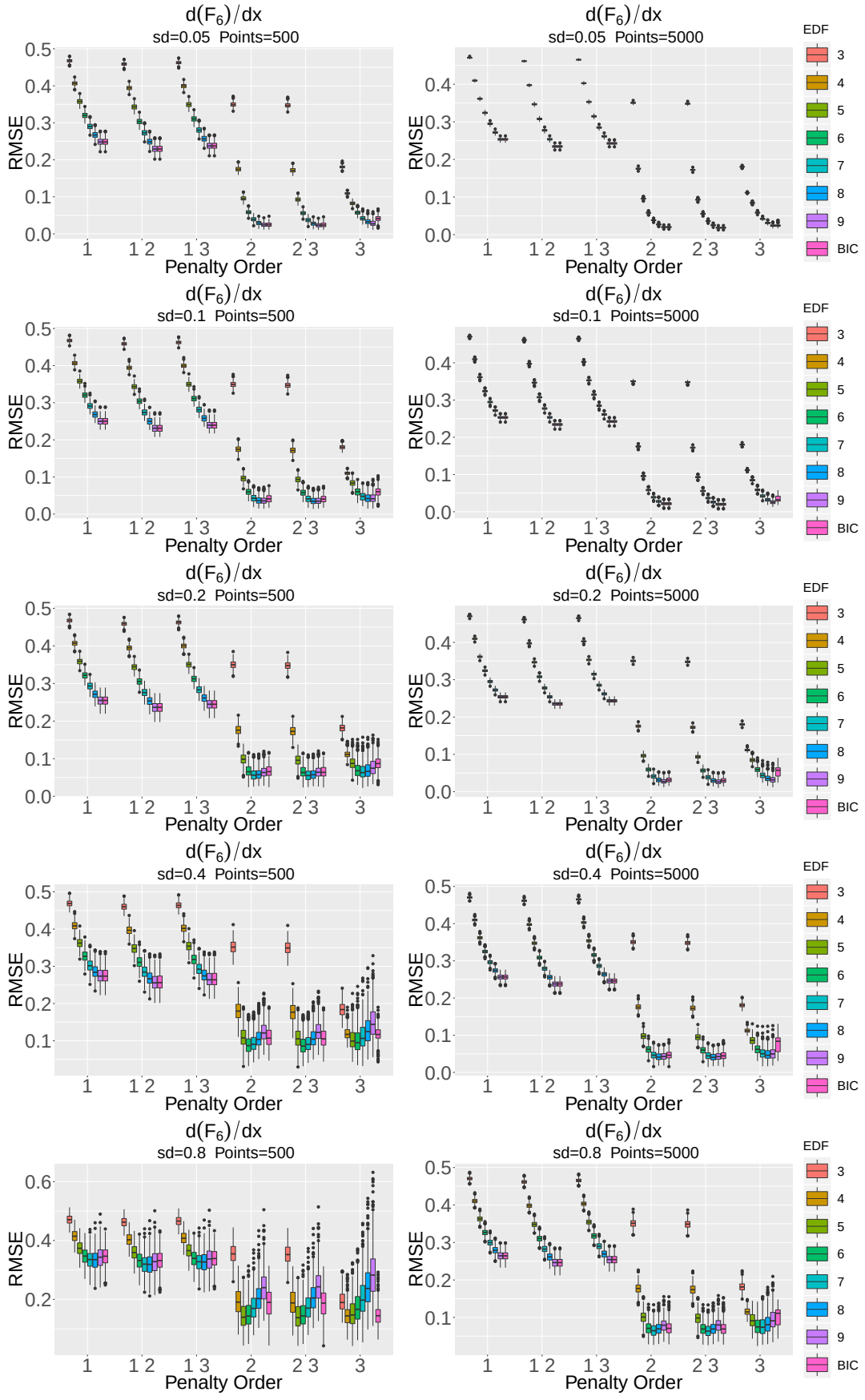


FIGURE A.6: Box plots of $\frac{d(F_6)}{dx}$ estimation RMSE for 500 and 5000 data points.

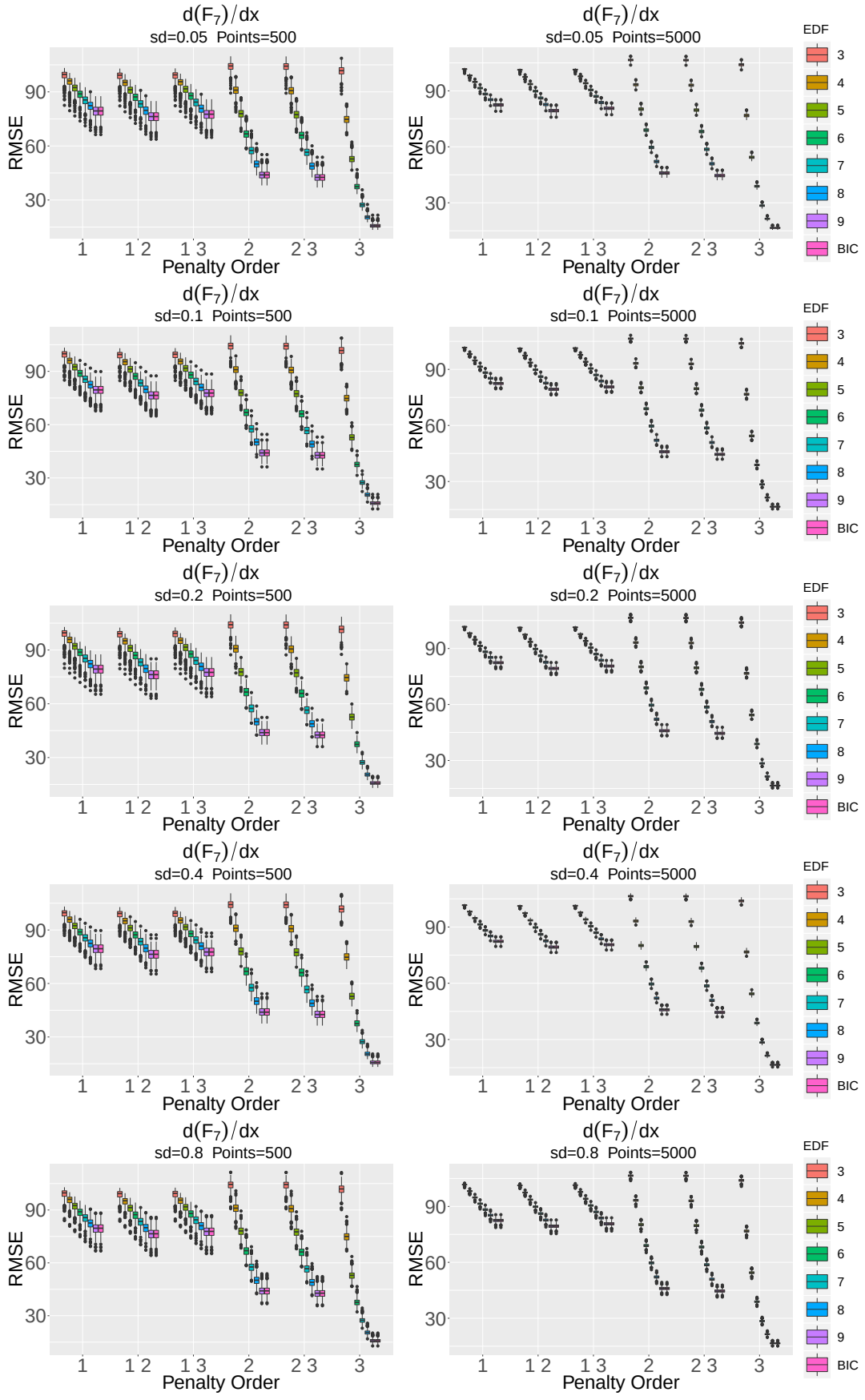


FIGURE A.7: Box plots of $\frac{d(F_7)}{dx}$ estimation RMSE for 500 and 5000 data points.

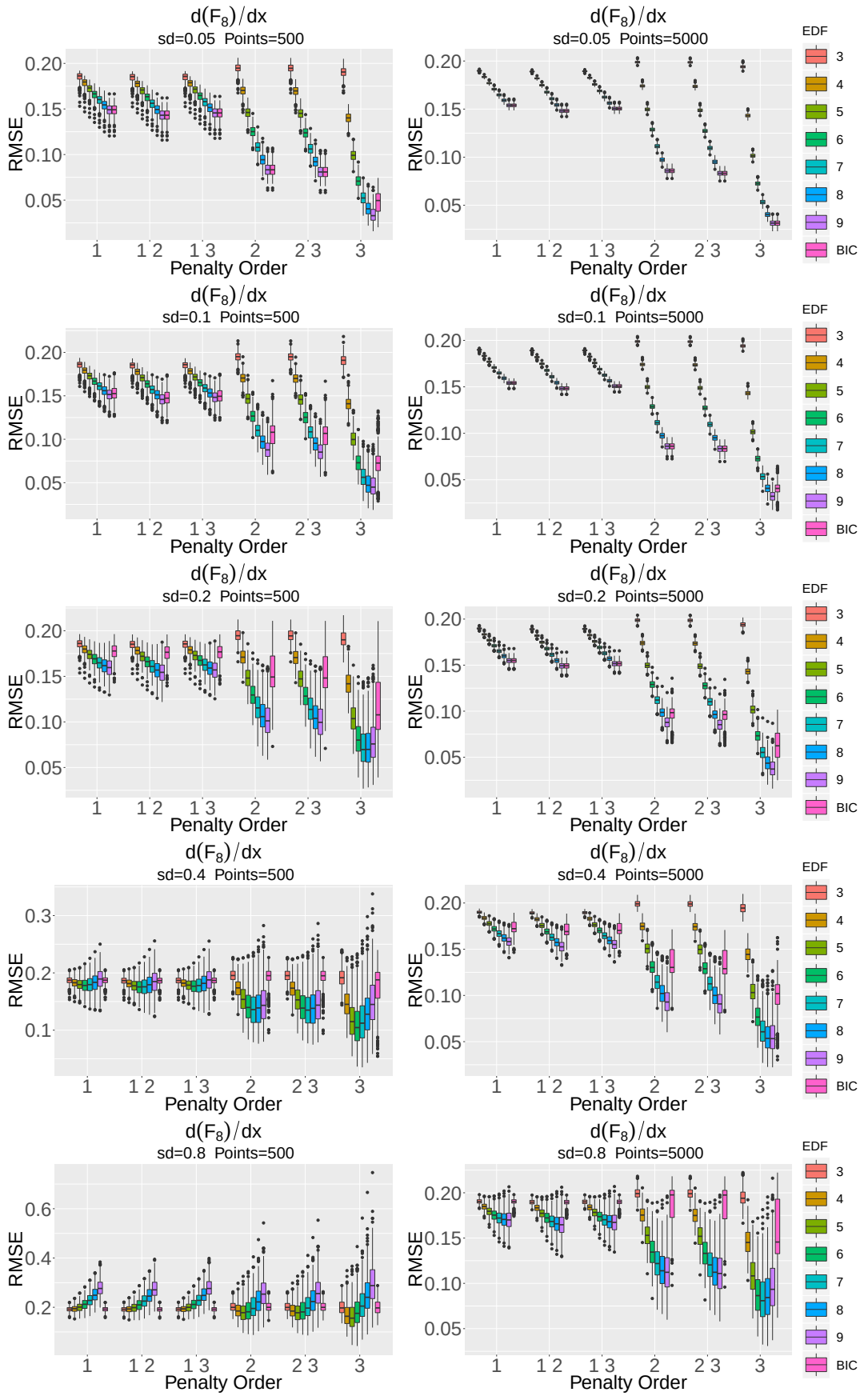


FIGURE A.8: Box plots of $\frac{d(F_8)}{dx}$ estimation RMSE for 500 and 5000 data points.

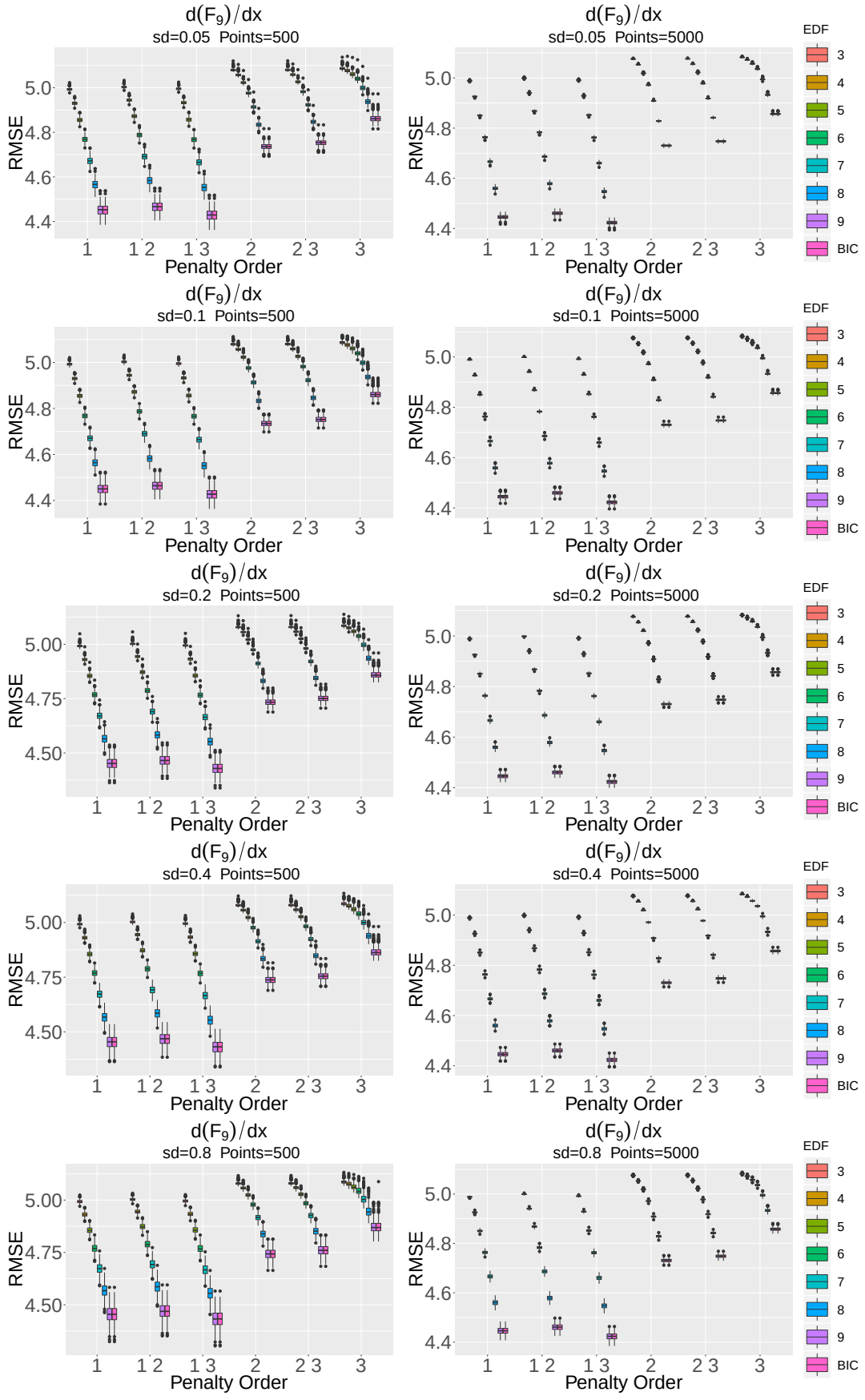


FIGURE A.9: Box plots of $\frac{d(F_9)}{dx}$ estimation RMSE for 500 and 5000 data points.

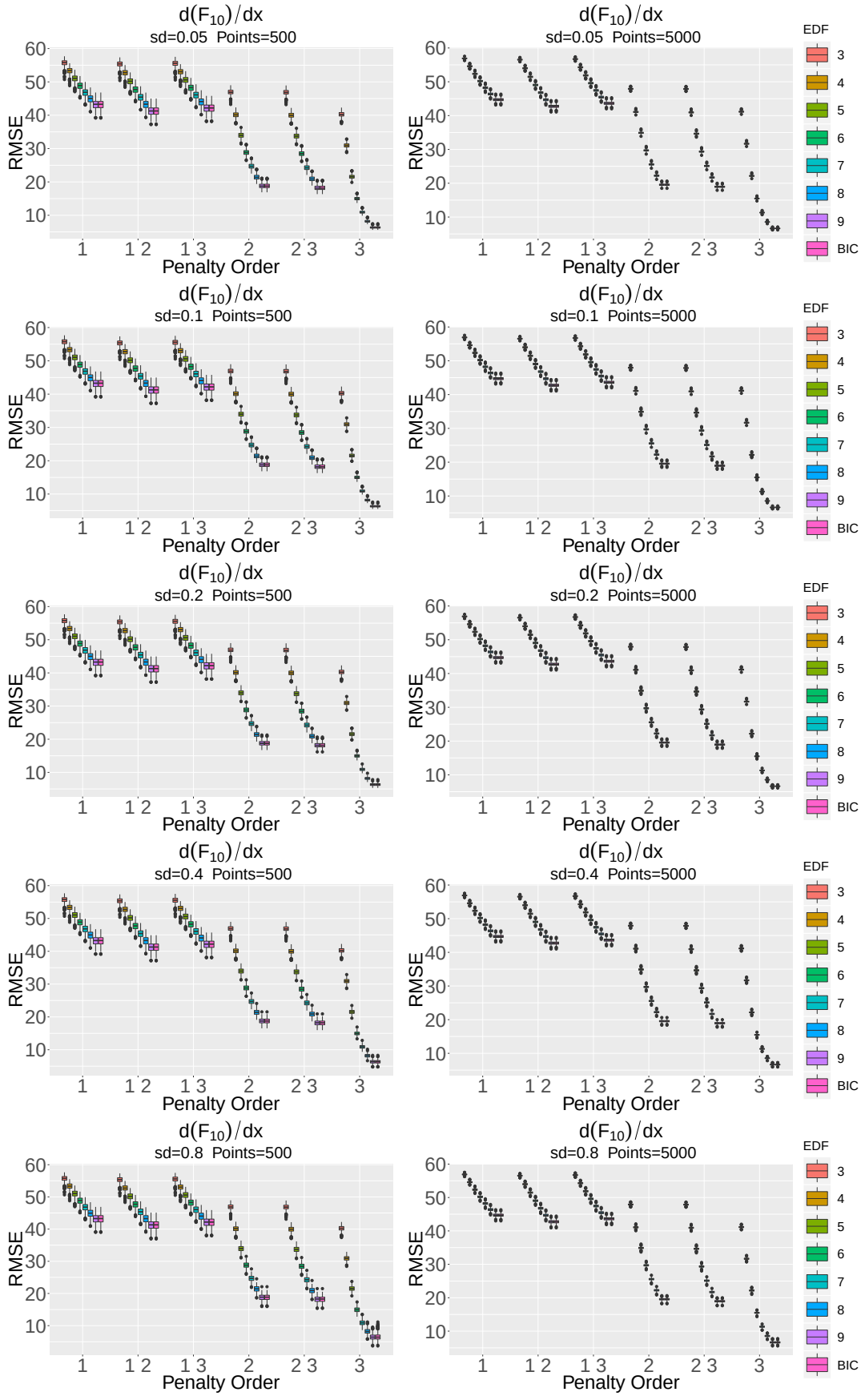


FIGURE A.10: Box plots of $\frac{d(F_{10})}{dx}$ estimation RMSE for 500 and 5000 data points.

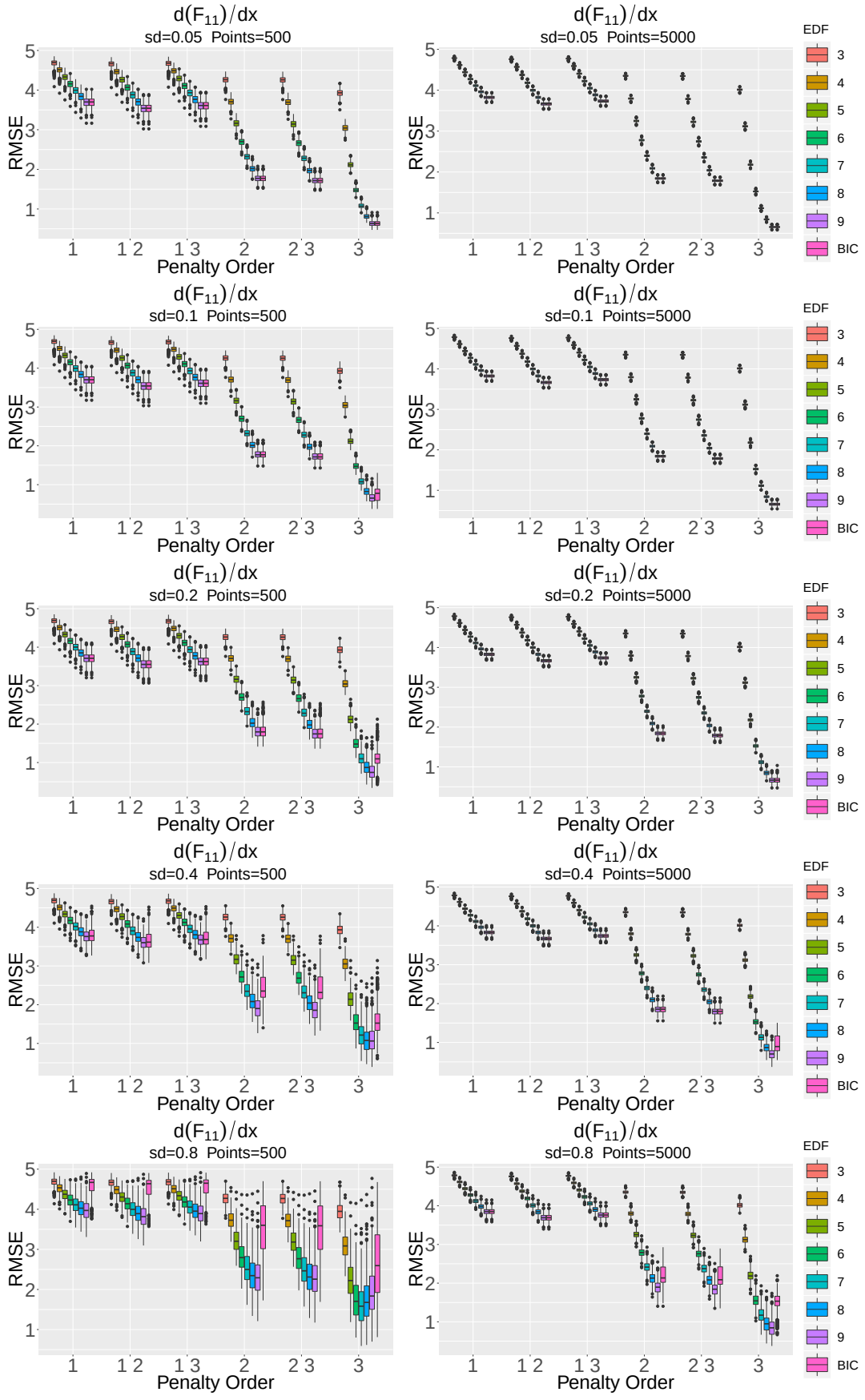


FIGURE A.11: Box plots of $\frac{d(F_{11})}{dx}$ estimation RMSE for 500 and 5000 data points.

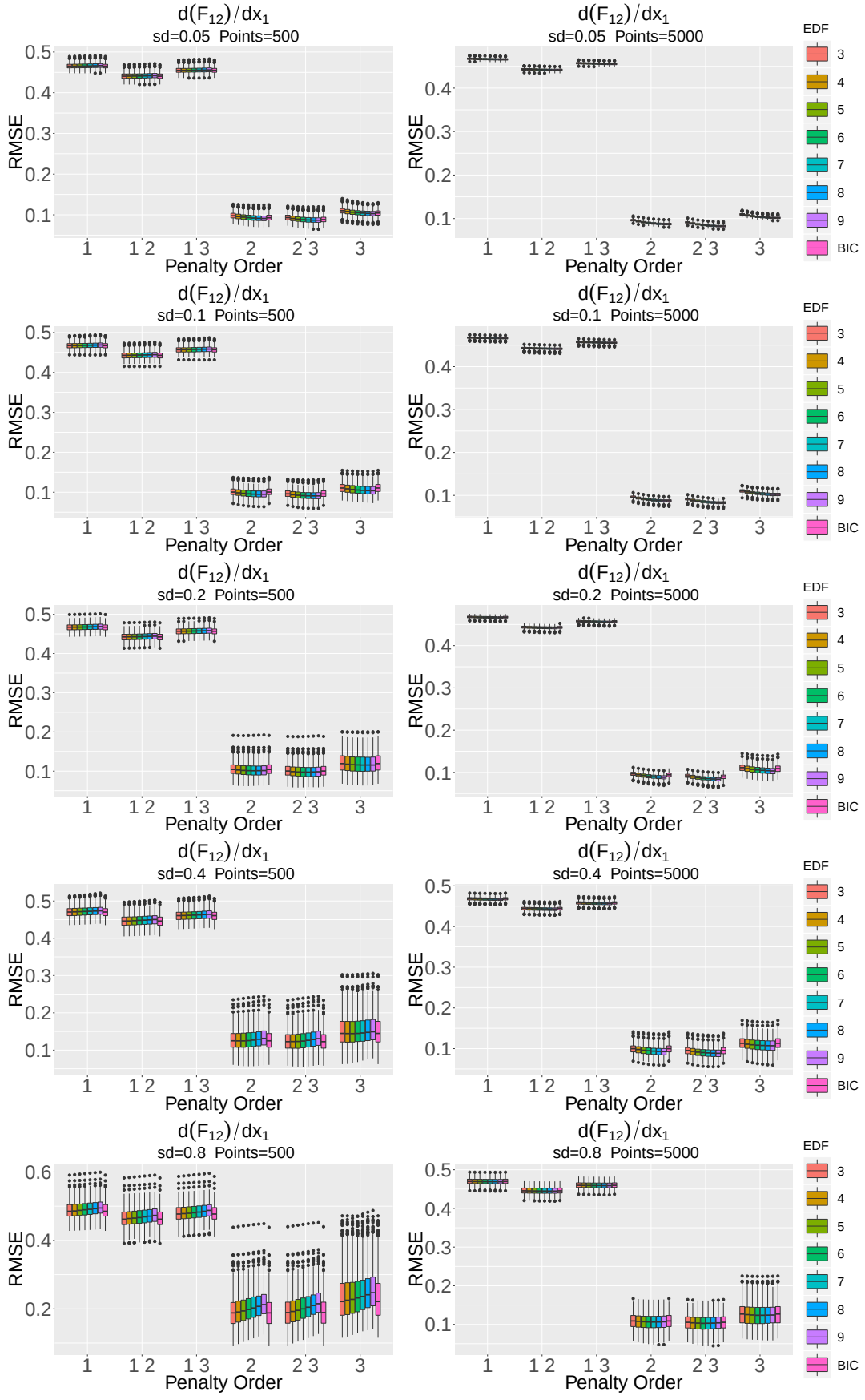


FIGURE A.12: Box plots of $\frac{\partial F_{12}}{\partial x_1}$ estimation RMSE for 500 and 5000 data points.

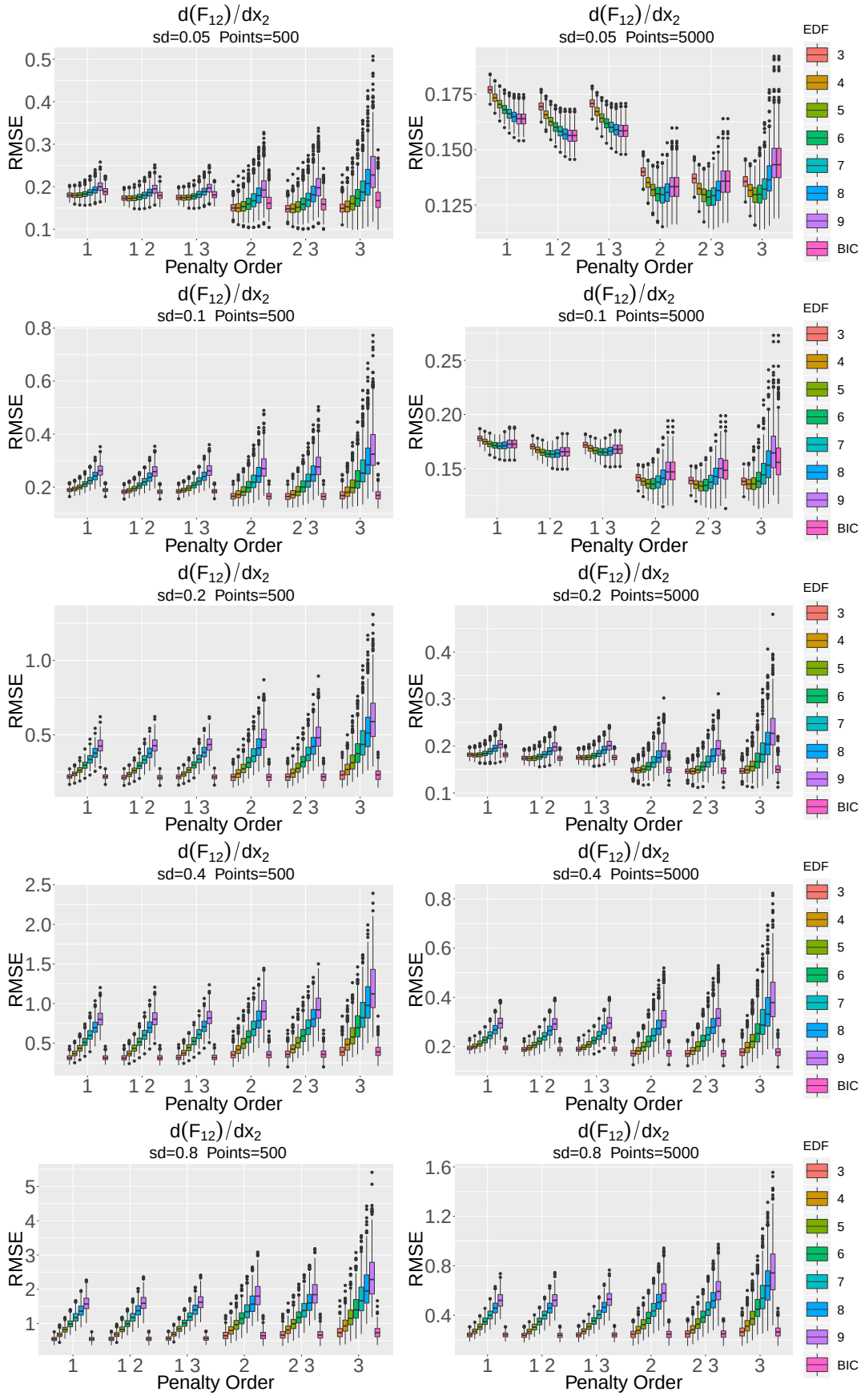


FIGURE A.13: Box plots of $\frac{\partial F_{12}}{\partial x_2}$ estimation RMSE for 500 and 5000 data points.

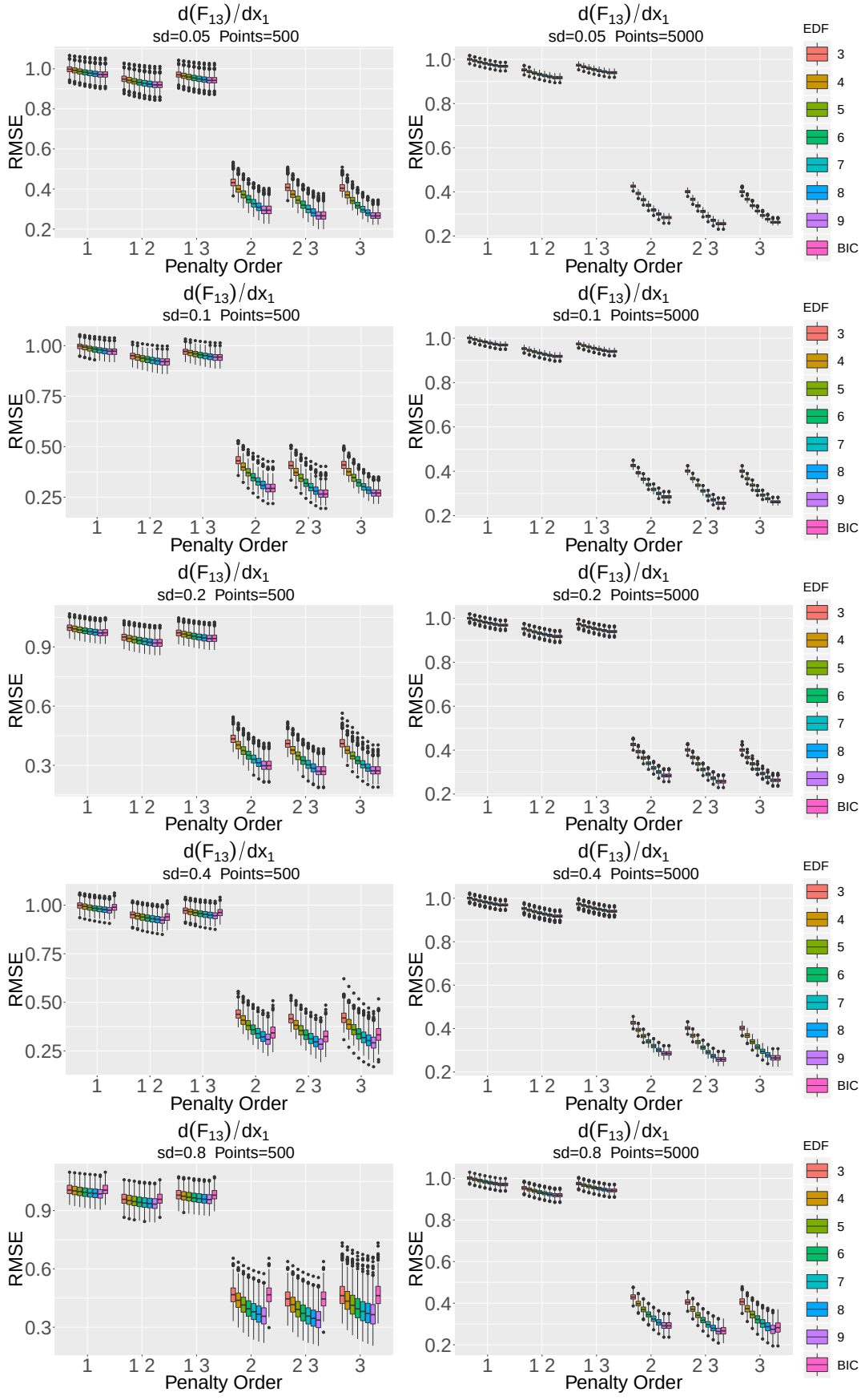


FIGURE A.14: Box plots of $\frac{\partial F_{13}}{\partial x_1}$ estimation RMSE for 500 and 5000 data points.

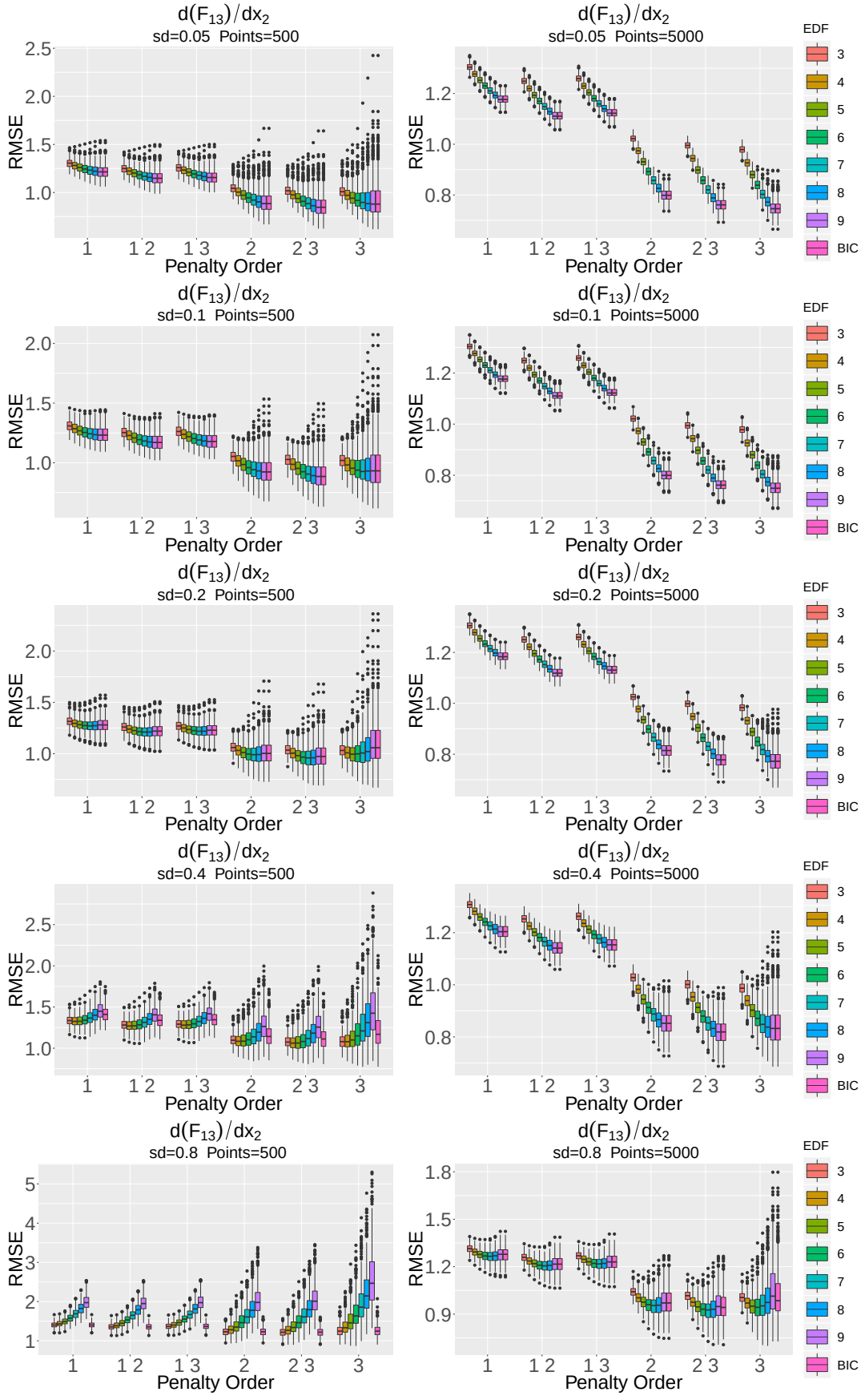


FIGURE A.15: Box plots of $\frac{\partial F_{13}}{\partial x_2}$ estimation RMSE for 500 and 5000 data points.

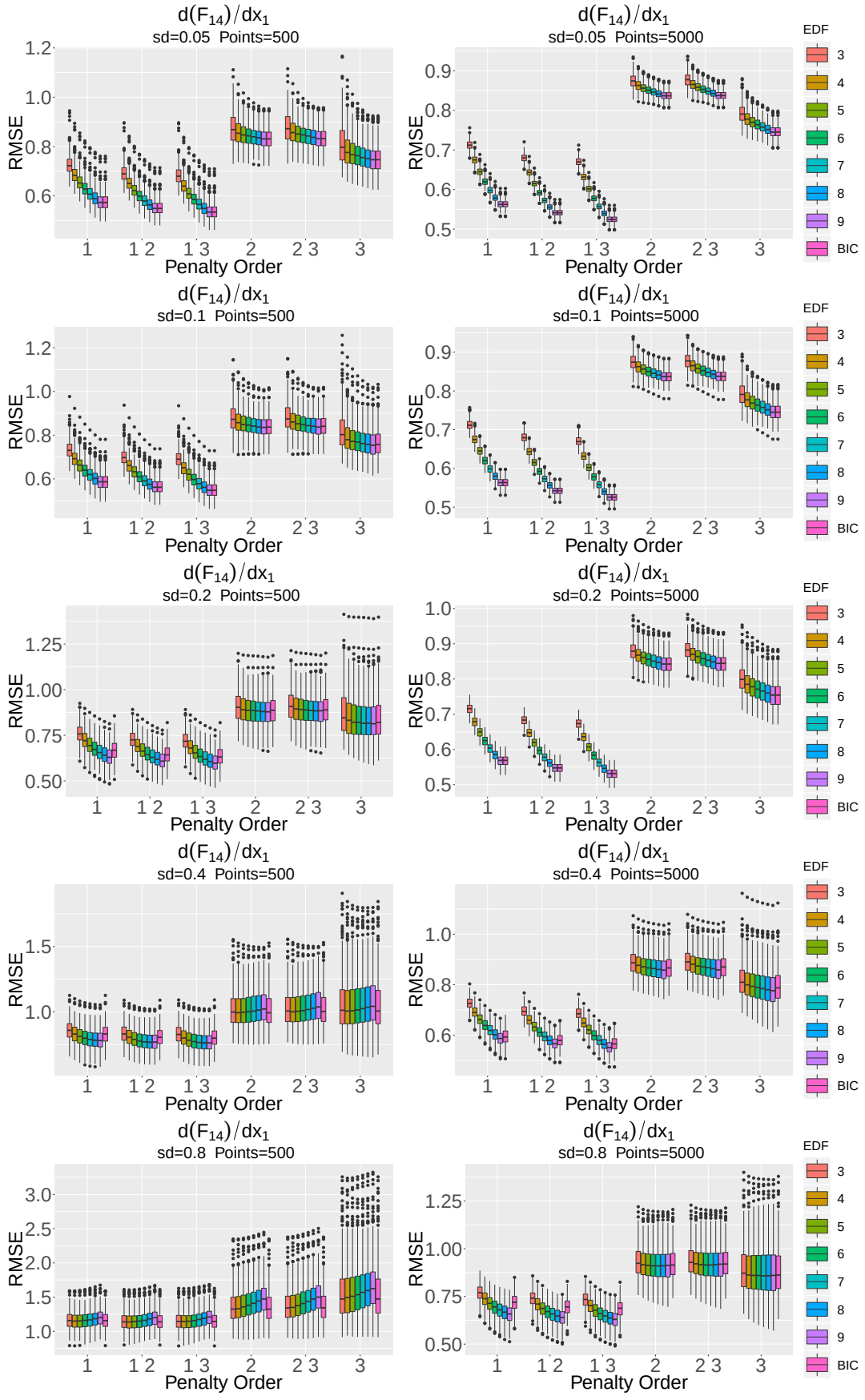


FIGURE A.16: Box plots of $\frac{\partial F_{14}}{\partial x_1}$ estimation RMSE for 500 and 5000 data points.

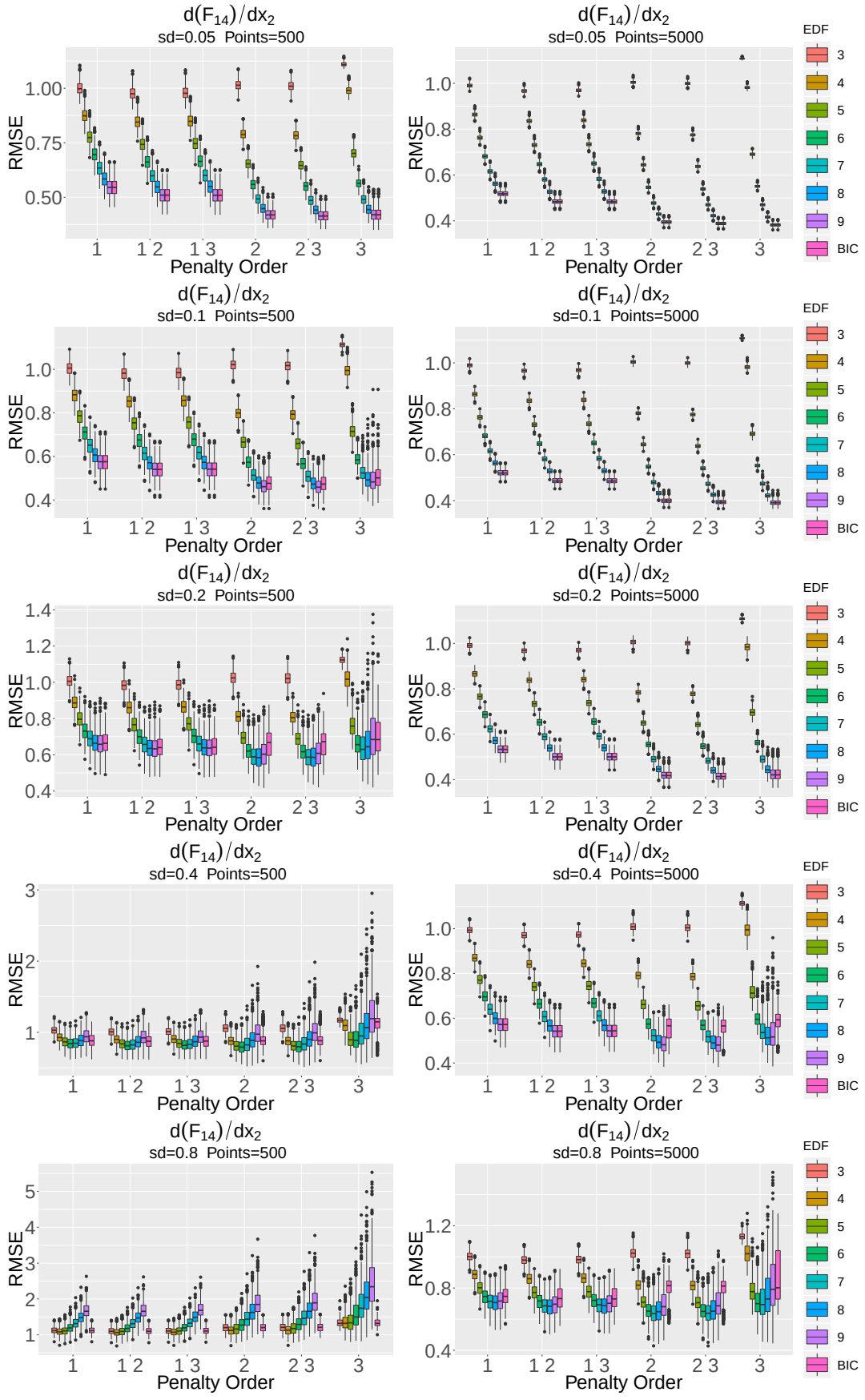


FIGURE A.17: Box plots of $\frac{\partial F_{14}}{\partial x_2}$ estimation RMSE for 500 and 5000 data points.

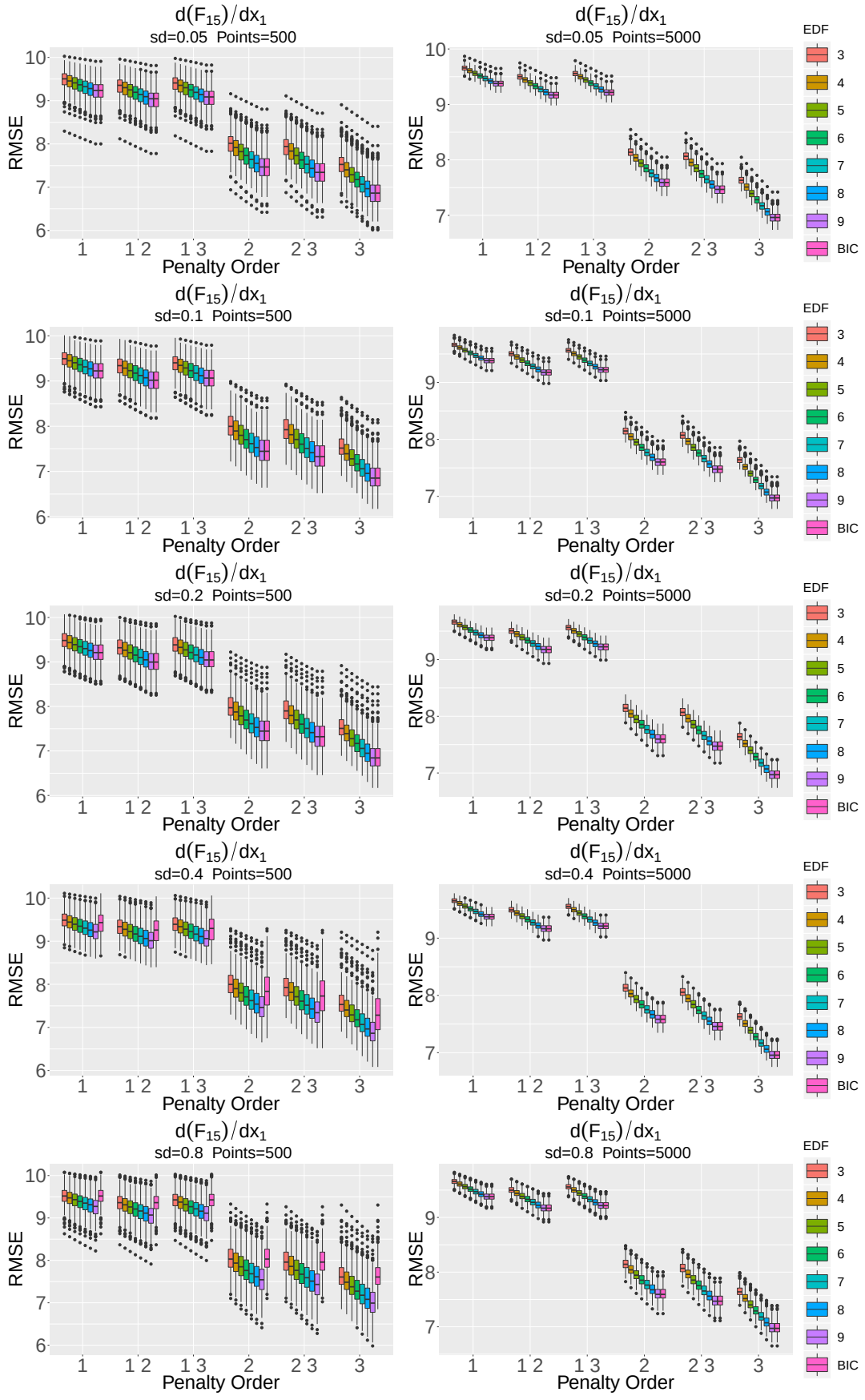


FIGURE A.18: Box plots of $\frac{\partial F_{15}}{\partial x_1}$ estimation RMSE for 500 and 5000 data points.

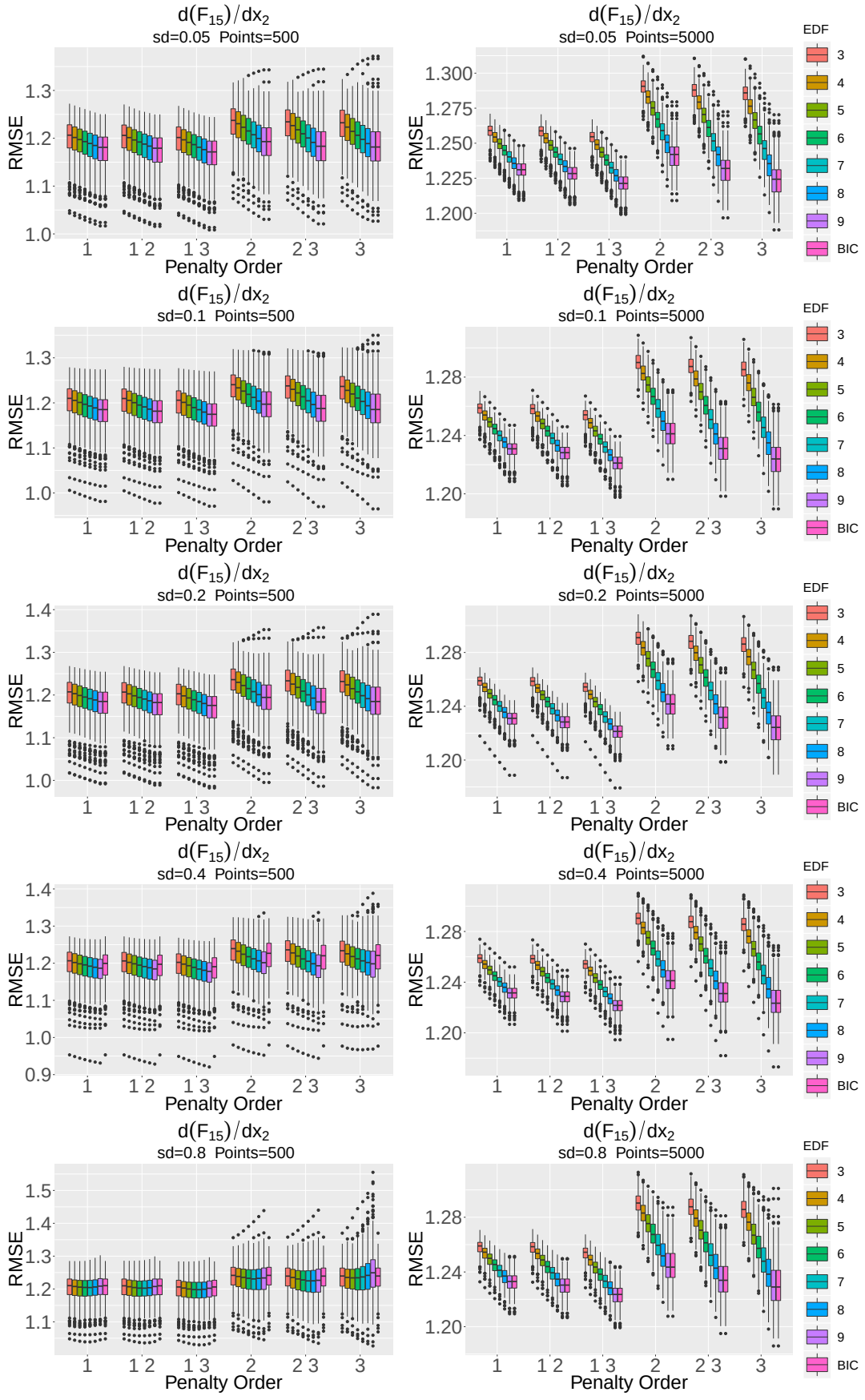


FIGURE A.19: Box plots of $\frac{\partial F_{15}}{\partial x_2}$ estimation RMSE for 500 and 5000 data points.

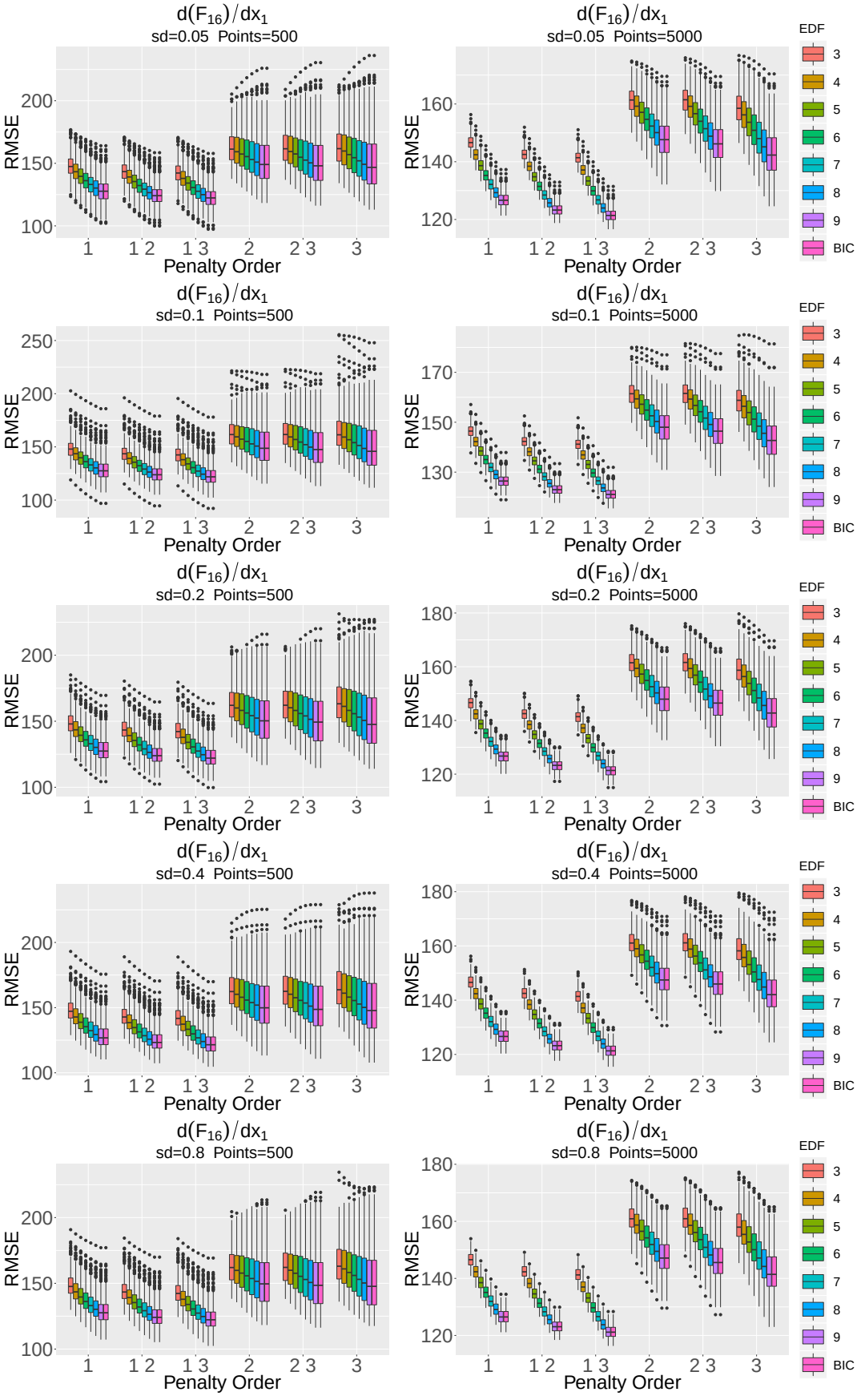


FIGURE A.20: Box plots of $\frac{\partial F_{16}}{\partial x_1}$ estimation RMSE for 500 and 5000 data points.

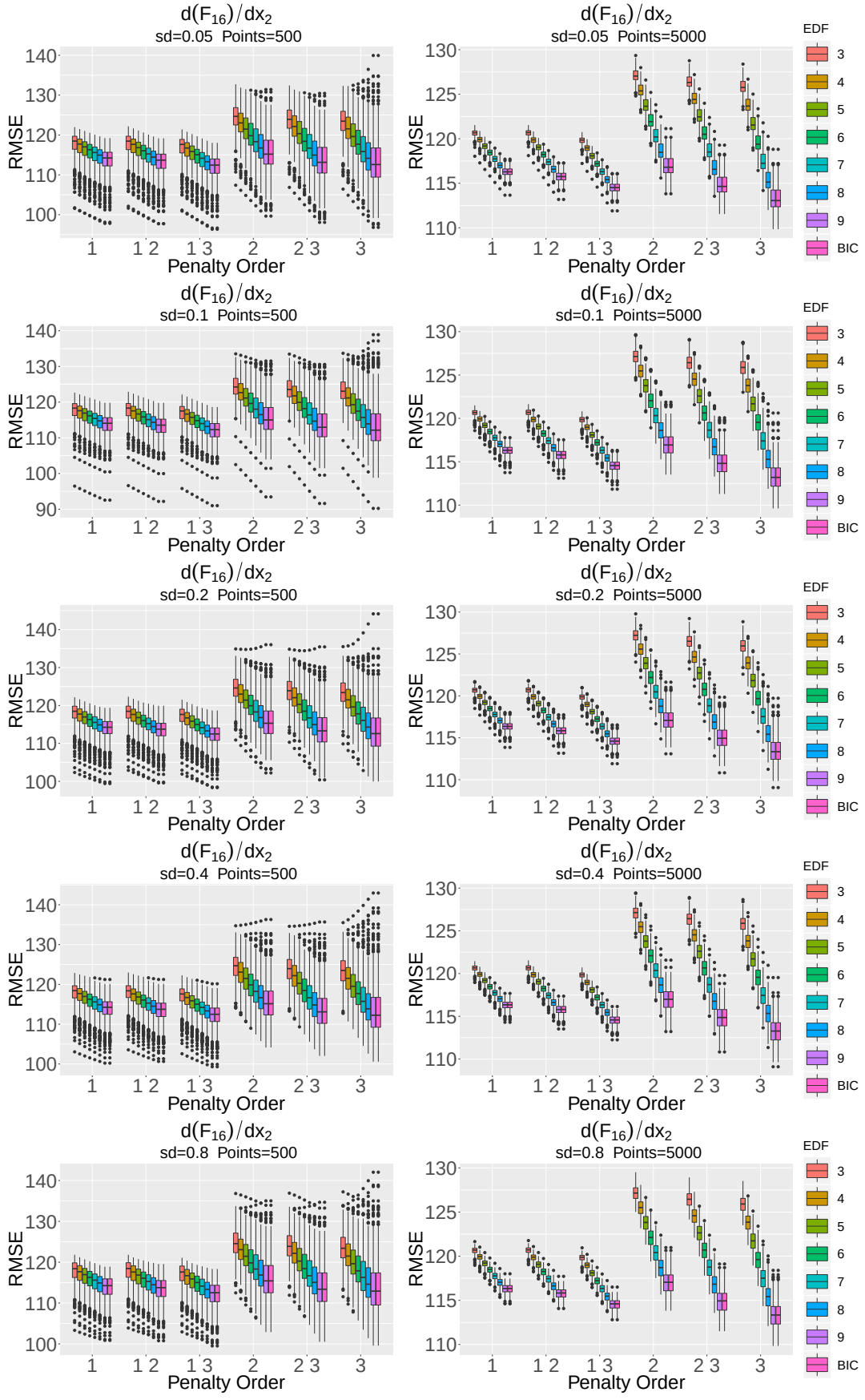


FIGURE A.21: Box plots of $\frac{\partial F_{16}}{\partial x_2}$ estimation RMSE for 500 and 5000 data points.

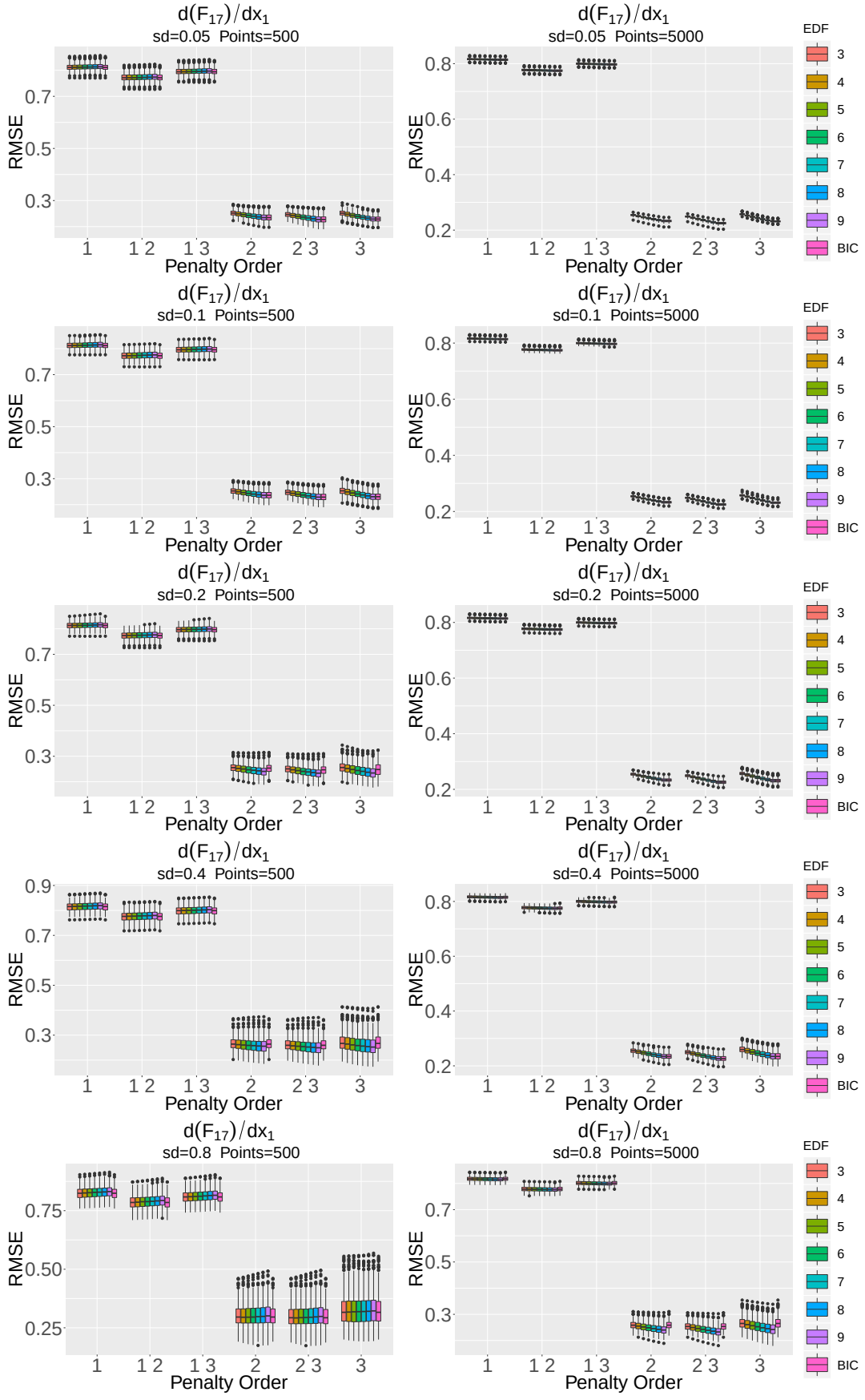


FIGURE A.22: Box plots of $\frac{\partial F_{17}}{\partial x_1}$ estimation RMSE for 500 and 5000 data points.

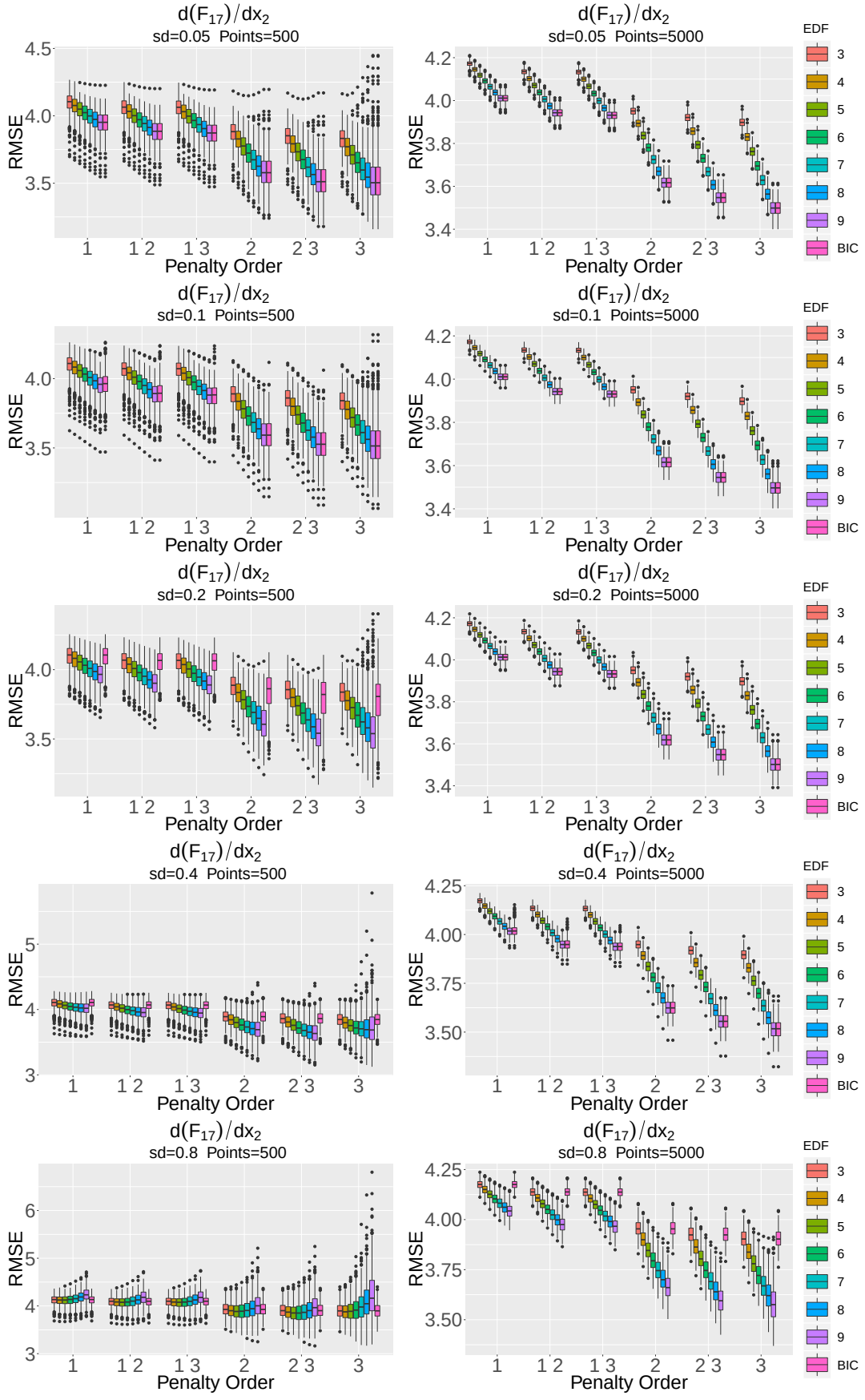


FIGURE A.23: Box plots of $\frac{\partial F_{17}}{\partial x_2}$ estimation RMSE for 500 and 5000 data points.

Function	σ	Penalty Order	Function	σ	Penalty Order
$\boldsymbol{F_1}$	0.02	p3	$\boldsymbol{F_2}$	0.02	p2
$\boldsymbol{F_1}$	0.05	p3	$\boldsymbol{F_2}$	0.05	p23
$\boldsymbol{F_1}$	0.10	p12	$\boldsymbol{F_2}$	0.10	p23
$\boldsymbol{F_1}$	0.20	p12	$\boldsymbol{F_2}$	0.20	p2
$\boldsymbol{F_1}$	0.40	p12	$\boldsymbol{F_2}$	0.40	p2
$\boldsymbol{F_1}$	0.80	p12	$\boldsymbol{F_2}$	0.80	p3
$\boldsymbol{F_3}$	0.02	p13	$\boldsymbol{F_4}$	0.02	p3
$\boldsymbol{F_3}$	0.05	p13	$\boldsymbol{F_4}$	0.05	p3
$\boldsymbol{F_3}$	0.10	p13	$\boldsymbol{F_4}$	0.10	p2
$\boldsymbol{F_3}$	0.20	p13	$\boldsymbol{F_4}$	0.20	p23
$\boldsymbol{F_3}$	0.40	p13	$\boldsymbol{F_4}$	0.40	p12
$\boldsymbol{F_3}$	0.80	p13	$\boldsymbol{F_4}$	0.80	p12
$\boldsymbol{F_5}$	0.02	p3	$\boldsymbol{F_6}$	0.02	p23
$\boldsymbol{F_5}$	0.05	p3	$\boldsymbol{F_6}$	0.05	p23
$\boldsymbol{F_5}$	0.10	p3	$\boldsymbol{F_6}$	0.10	p23
$\boldsymbol{F_5}$	0.20	p12	$\boldsymbol{F_6}$	0.20	p23
$\boldsymbol{F_5}$	0.40	p12	$\boldsymbol{F_6}$	0.40	p23
$\boldsymbol{F_5}$	0.80	p12	$\boldsymbol{F_6}$	0.80	p3
$\boldsymbol{F_7}$	0.02	p3	$\boldsymbol{F_8}$	0.02	p3
$\boldsymbol{F_7}$	0.05	p3	$\boldsymbol{F_8}$	0.05	p3
$\boldsymbol{F_7}$	0.10	p3	$\boldsymbol{F_8}$	0.10	p3
$\boldsymbol{F_7}$	0.20	p3	$\boldsymbol{F_8}$	0.20	p3
$\boldsymbol{F_7}$	0.40	p3	$\boldsymbol{F_8}$	0.40	p12
$\boldsymbol{F_7}$	0.80	p3	$\boldsymbol{F_8}$	0.80	p12
$\boldsymbol{F_9}$	0.02	p13	$\boldsymbol{F_{10}}$	0.02	p3
$\boldsymbol{F_9}$	0.05	p13	$\boldsymbol{F_{10}}$	0.05	p3
$\boldsymbol{F_9}$	0.10	p13	$\boldsymbol{F_{10}}$	0.10	p3
$\boldsymbol{F_9}$	0.20	p13	$\boldsymbol{F_{10}}$	0.20	p3
$\boldsymbol{F_9}$	0.40	p13	$\boldsymbol{F_{10}}$	0.40	p3
$\boldsymbol{F_9}$	0.80	p13	$\boldsymbol{F_{10}}$	0.80	p3
$\boldsymbol{F_{11}}$	0.02	p3	*	*	*
$\boldsymbol{F_{11}}$	0.05	p3	*	*	*
$\boldsymbol{F_{11}}$	0.10	p3	*	*	*
$\boldsymbol{F_{11}}$	0.20	p3	*	*	*
$\boldsymbol{F_{11}}$	0.40	p3	*	*	*
$\boldsymbol{F_{11}}$	0.80	p3	*	*	*

TABLE A.1: The table above depicts the penalty orders yielding the smallest RMSE for derivatives when using the optimal degrees of freedom for 500 points.

Function	σ	Penalty Order	Function	σ	Penalty Order
$\boldsymbol{F_1}$	0.02	p3	$\boldsymbol{F_2}$	0.02	p2
$\boldsymbol{F_1}$	0.05	p3	$\boldsymbol{F_2}$	0.05	p2
$\boldsymbol{F_1}$	0.10	p3	$\boldsymbol{F_2}$	0.10	p2
$\boldsymbol{F_1}$	0.20	p3	$\boldsymbol{F_2}$	0.20	p2
$\boldsymbol{F_1}$	0.40	p12	$\boldsymbol{F_2}$	0.40	p2
$\boldsymbol{F_1}$	0.80	p12	$\boldsymbol{F_2}$	0.80	p2
$\boldsymbol{F_3}$	0.02	p13	$\boldsymbol{F_4}$	0.02	p3
$\boldsymbol{F_3}$	0.05	p13	$\boldsymbol{F_4}$	0.05	p3
$\boldsymbol{F_3}$	0.10	p13	$\boldsymbol{F_4}$	0.10	p3
$\boldsymbol{F_3}$	0.20	p13	$\boldsymbol{F_4}$	0.20	p2
$\boldsymbol{F_3}$	0.40	p13	$\boldsymbol{F_4}$	0.40	p2
$\boldsymbol{F_3}$	0.80	p13	$\boldsymbol{F_4}$	0.80	p23
$\boldsymbol{F_5}$	0.02	p3	$\boldsymbol{F_6}$	0.02	p23
$\boldsymbol{F_5}$	0.05	p3	$\boldsymbol{F_6}$	0.05	p23
$\boldsymbol{F_5}$	0.10	p3	$\boldsymbol{F_6}$	0.10	p23
$\boldsymbol{F_5}$	0.20	p3	$\boldsymbol{F_6}$	0.20	p23
$\boldsymbol{F_5}$	0.40	p12	$\boldsymbol{F_6}$	0.40	p23
$\boldsymbol{F_5}$	0.80	p12	$\boldsymbol{F_6}$	0.80	p23
$\boldsymbol{F_7}$	0.02	p3	$\boldsymbol{F_8}$	0.02	p3
$\boldsymbol{F_7}$	0.05	p3	$\boldsymbol{F_8}$	0.05	p3
$\boldsymbol{F_7}$	0.10	p3	$\boldsymbol{F_8}$	0.10	p3
$\boldsymbol{F_7}$	0.20	p3	$\boldsymbol{F_8}$	0.20	p3
$\boldsymbol{F_7}$	0.40	p3	$\boldsymbol{F_8}$	0.40	p3
$\boldsymbol{F_7}$	0.80	p3	$\boldsymbol{F_8}$	0.80	p3
$\boldsymbol{F_9}$	0.02	p13	$\boldsymbol{F_{10}}$	0.02	p3
$\boldsymbol{F_9}$	0.05	p13	$\boldsymbol{F_{10}}$	0.05	p3
$\boldsymbol{F_9}$	0.10	p13	$\boldsymbol{F_{10}}$	0.10	p3
$\boldsymbol{F_9}$	0.20	p13	$\boldsymbol{F_{10}}$	0.20	p3
$\boldsymbol{F_9}$	0.40	p13	$\boldsymbol{F_{10}}$	0.40	p3
$\boldsymbol{F_9}$	0.80	p13	$\boldsymbol{F_{10}}$	0.80	p3
$\boldsymbol{F_{11}}$	0.02	p3	*	*	*
$\boldsymbol{F_{11}}$	0.05	p3	*	*	*
$\boldsymbol{F_{11}}$	0.10	p3	*	*	*
$\boldsymbol{F_{11}}$	0.20	p3	*	*	*
$\boldsymbol{F_{11}}$	0.40	p3	*	*	*
$\boldsymbol{F_{11}}$	0.80	p3	*	*	*

TABLE A.2: The table above depicts the penalty orders yielding the smallest RMSE for derivatives when using the optimal degrees of freedom for 5000 points.

Function	σ	Penalty Order	Function	σ	Penalty Order
$\boldsymbol{F_{12}}$ wrt x_1	0.02	p23	$\boldsymbol{F_{12}}$ wrt x_2	0.02	p2
$\boldsymbol{F_{12}}$ wrt x_1	0.05	p23	$\boldsymbol{F_{12}}$ wrt x_2	0.05	p23
$\boldsymbol{F_{12}}$ wrt x_1	0.1	p2	$\boldsymbol{F_{12}}$ wrt x_2	0.1	p23
$\boldsymbol{F_{12}}$ wrt x_1	0.2	p23	$\boldsymbol{F_{12}}$ wrt x_2	0.2	p12
$\boldsymbol{F_{12}}$ wrt x_1	0.4	p23	$\boldsymbol{F_{12}}$ wrt x_2	0.4	p12
$\boldsymbol{F_{12}}$ wrt x_1	0.8	p3	$\boldsymbol{F_{12}}$ wrt x_2	0.8	p1
$\boldsymbol{F_{13}}$ wrt x_1	0.02	p3	$\boldsymbol{F_{13}}$ wrt x_2	0.02	p23
$\boldsymbol{F_{13}}$ wrt x_1	0.05	p3	$\boldsymbol{F_{13}}$ wrt x_2	0.05	p23
$\boldsymbol{F_{13}}$ wrt x_1	0.1	p3	$\boldsymbol{F_{13}}$ wrt x_2	0.1	p23
$\boldsymbol{F_{13}}$ wrt x_1	0.2	p3	$\boldsymbol{F_{13}}$ wrt x_2	0.2	p23
$\boldsymbol{F_{13}}$ wrt x_1	0.4	p23	$\boldsymbol{F_{13}}$ wrt x_2	0.4	p23
$\boldsymbol{F_{13}}$ wrt x_1	0.8	p23	$\boldsymbol{F_{13}}$ wrt x_2	0.8	p23
$\boldsymbol{F_{14}}$ wrt x_1	0.02	p13	$\boldsymbol{F_{14}}$ wrt x_2	0.02	p3
$\boldsymbol{F_{14}}$ wrt x_1	0.05	p13	$\boldsymbol{F_{14}}$ wrt x_2	0.05	p23
$\boldsymbol{F_{14}}$ wrt x_1	0.1	p13	$\boldsymbol{F_{14}}$ wrt x_2	0.1	p23
$\boldsymbol{F_{14}}$ wrt x_1	0.2	p13	$\boldsymbol{F_{14}}$ wrt x_2	0.2	p12
$\boldsymbol{F_{14}}$ wrt x_1	0.4	p13	$\boldsymbol{F_{14}}$ wrt x_2	0.4	p12
$\boldsymbol{F_{14}}$ wrt x_1	0.8	p12	$\boldsymbol{F_{14}}$ wrt x_2	0.8	p12
$\boldsymbol{F_{15}}$ wrt x_1	0.02	p3	$\boldsymbol{F_{15}}$ wrt x_2	0.02	p13
$\boldsymbol{F_{15}}$ wrt x_1	0.05	p3	$\boldsymbol{F_{15}}$ wrt x_2	0.05	p13
$\boldsymbol{F_{15}}$ wrt x_1	0.1	p3	$\boldsymbol{F_{15}}$ wrt x_2	0.1	p13
$\boldsymbol{F_{15}}$ wrt x_1	0.2	p3	$\boldsymbol{F_{15}}$ wrt x_2	0.2	p13
$\boldsymbol{F_{15}}$ wrt x_1	0.4	p3	$\boldsymbol{F_{15}}$ wrt x_2	0.4	p13
$\boldsymbol{F_{15}}$ wrt x_1	0.8	p3	$\boldsymbol{F_{15}}$ wrt x_2	0.8	p13
$\boldsymbol{F_{16}}$ wrt x_1	0.02	p13	$\boldsymbol{F_{16}}$ wrt x_2	0.02	p3
$\boldsymbol{F_{16}}$ wrt x_1	0.05	p13	$\boldsymbol{F_{16}}$ wrt x_2	0.05	p13
$\boldsymbol{F_{16}}$ wrt x_1	0.1	p13	$\boldsymbol{F_{16}}$ wrt x_2	0.1	p3
$\boldsymbol{F_{16}}$ wrt x_1	0.2	p13	$\boldsymbol{F_{16}}$ wrt x_2	0.2	p13
$\boldsymbol{F_{16}}$ wrt x_1	0.4	p13	$\boldsymbol{F_{16}}$ wrt x_2	0.4	p3
$\boldsymbol{F_{16}}$ wrt x_1	0.8	p13	$\boldsymbol{F_{16}}$ wrt x_2	0.8	p13
$\boldsymbol{F_{17}}$ wrt x_1	0.02	p23	$\boldsymbol{F_{17}}$ wrt x_2	0.02	p3
$\boldsymbol{F_{17}}$ wrt x_1	0.05	p23	$\boldsymbol{F_{17}}$ wrt x_2	0.05	p3
$\boldsymbol{F_{17}}$ wrt x_1	0.1	p23	$\boldsymbol{F_{17}}$ wrt x_2	0.1	p3
$\boldsymbol{F_{17}}$ wrt x_1	0.2	p23	$\boldsymbol{F_{17}}$ wrt x_2	0.2	p3
$\boldsymbol{F_{17}}$ wrt x_1	0.4	p23	$\boldsymbol{F_{17}}$ wrt x_2	0.4	p3
$\boldsymbol{F_{17}}$ wrt x_1	0.8	p23	$\boldsymbol{F_{17}}$ wrt x_2	0.8	p3

TABLE A.3: The table above depicts the penalty orders yielding the smallest RMSE for partial derivative when using the optimal degrees of freedom for 500 points.

Function	σ	Penalty Order	Function	σ	Penalty Order
$\boldsymbol{F_{12}}$ wrt x_1	0.02	p23	$\boldsymbol{F_{12}}$ wrt x_2	0.02	p2
$\boldsymbol{F_{12}}$ wrt x_1	0.05	p23	$\boldsymbol{F_{12}}$ wrt x_2	0.05	p2
$\boldsymbol{F_{12}}$ wrt x_1	0.1	p23	$\boldsymbol{F_{12}}$ wrt x_2	0.1	p2
$\boldsymbol{F_{12}}$ wrt x_1	0.2	p23	$\boldsymbol{F_{12}}$ wrt x_2	0.2	p23
$\boldsymbol{F_{12}}$ wrt x_1	0.4	p23	$\boldsymbol{F_{12}}$ wrt x_2	0.4	p23
$\boldsymbol{F_{12}}$ wrt x_1	0.8	p2	$\boldsymbol{F_{12}}$ wrt x_2	0.8	p12
$\boldsymbol{F_{13}}$ wrt x_1	0.02	p3	$\boldsymbol{F_{13}}$ wrt x_2	0.02	p3
$\boldsymbol{F_{13}}$ wrt x_1	0.05	p3	$\boldsymbol{F_{13}}$ wrt x_2	0.05	p3
$\boldsymbol{F_{13}}$ wrt x_1	0.1	p3	$\boldsymbol{F_{13}}$ wrt x_2	0.1	p3
$\boldsymbol{F_{13}}$ wrt x_1	0.2	p3	$\boldsymbol{F_{13}}$ wrt x_2	0.2	p3
$\boldsymbol{F_{13}}$ wrt x_1	0.4	p3	$\boldsymbol{F_{13}}$ wrt x_2	0.4	p23
$\boldsymbol{F_{13}}$ wrt x_1	0.8	p23	$\boldsymbol{F_{13}}$ wrt x_2	0.8	p23
$\boldsymbol{F_{14}}$ wrt x_1	0.02	p13	$\boldsymbol{F_{14}}$ wrt x_2	0.02	p3
$\boldsymbol{F_{14}}$ wrt x_1	0.05	p13	$\boldsymbol{F_{14}}$ wrt x_2	0.05	p3
$\boldsymbol{F_{14}}$ wrt x_1	0.1	p13	$\boldsymbol{F_{14}}$ wrt x_2	0.1	p3
$\boldsymbol{F_{14}}$ wrt x_1	0.2	p13	$\boldsymbol{F_{14}}$ wrt x_2	0.2	p23
$\boldsymbol{F_{14}}$ wrt x_1	0.4	p13	$\boldsymbol{F_{14}}$ wrt x_2	0.4	p12
$\boldsymbol{F_{14}}$ wrt x_1	0.8	p13	$\boldsymbol{F_{14}}$ wrt x_2	0.8	p13
$\boldsymbol{F_{15}}$ wrt x_1	0.02	p3	$\boldsymbol{F_{15}}$ wrt x_2	0.02	p13
$\boldsymbol{F_{15}}$ wrt x_1	0.05	p3	$\boldsymbol{F_{15}}$ wrt x_2	0.05	p13
$\boldsymbol{F_{15}}$ wrt x_1	0.1	p3	$\boldsymbol{F_{15}}$ wrt x_2	0.1	p13
$\boldsymbol{F_{15}}$ wrt x_1	0.2	p3	$\boldsymbol{F_{15}}$ wrt x_2	0.2	p13
$\boldsymbol{F_{15}}$ wrt x_1	0.4	p3	$\boldsymbol{F_{15}}$ wrt x_2	0.4	p13
$\boldsymbol{F_{15}}$ wrt x_1	0.8	p3	$\boldsymbol{F_{15}}$ wrt x_2	0.8	p13
$\boldsymbol{F_{16}}$ wrt x_1	0.02	p13	$\boldsymbol{F_{16}}$ wrt x_2	0.02	p3
$\boldsymbol{F_{16}}$ wrt x_1	0.05	p13	$\boldsymbol{F_{16}}$ wrt x_2	0.05	p3
$\boldsymbol{F_{16}}$ wrt x_1	0.1	p13	$\boldsymbol{F_{16}}$ wrt x_2	0.1	p3
$\boldsymbol{F_{16}}$ wrt x_1	0.2	p13	$\boldsymbol{F_{16}}$ wrt x_2	0.2	p3
$\boldsymbol{F_{16}}$ wrt x_1	0.4	p13	$\boldsymbol{F_{16}}$ wrt x_2	0.4	p3
$\boldsymbol{F_{16}}$ wrt x_1	0.8	p13	$\boldsymbol{F_{16}}$ wrt x_2	0.8	p3
$\boldsymbol{F_{17}}$ wrt x_1	0.02	p23	$\boldsymbol{F_{17}}$ wrt x_2	0.02	p3
$\boldsymbol{F_{17}}$ wrt x_1	0.05	p23	$\boldsymbol{F_{17}}$ wrt x_2	0.05	p3
$\boldsymbol{F_{17}}$ wrt x_1	0.1	p23	$\boldsymbol{F_{17}}$ wrt x_2	0.1	p3
$\boldsymbol{F_{17}}$ wrt x_1	0.2	p23	$\boldsymbol{F_{17}}$ wrt x_2	0.2	p3
$\boldsymbol{F_{17}}$ wrt x_1	0.4	p23	$\boldsymbol{F_{17}}$ wrt x_2	0.4	p3
$\boldsymbol{F_{17}}$ wrt x_1	0.8	p23	$\boldsymbol{F_{17}}$ wrt x_2	0.8	p3

TABLE A.4: The table above depicts the penalty orders yielding the smallest RMSE for partial derivative when using the optimal degrees of freedom for 5000 points.

Bibliography

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle–in: Second international symposium on information theory (eds) bn petrov, f. csaki. *BNPBF Csaki. Budapest: Academiai Kiado*.

Aldrin, M. (2006). Improved predictions penalizing both slope and curvature in additive models. *Computational statistics & data analysis 50*(2), 267–284.

Baxter, J. M., I. L. Boyd, M. Cox, A. Donald, S. Malcolm, H. Miles, B. Miller, and C. Moffat (2011). Scotlands marine atlas: Information for the national marine plan. *Marine Scotland, Edinburgh*, 191.

Belitz, C. and S. Lang (2008). Simultaneous selection of variables and smoothing parameters in structured additive regression models. *Computational Statistics & Data Analysis 53*(1), 61–81.

Bowman, A. W. and A. Azzalini (1997). *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*, Volume 18. OUP Oxford.

Chakrabarti, A. and J. K. Ghosh (2011). Aic, bic and recent advances in model selection. In *Philosophy of Statistics*, pp. 583–605. Elsevier.

Charnigo, R., B. Hall, and C. Srinivasan (2011). A generalized c p criterion for derivative estimation. *Technometrics 53*(3), 238–253.

De Boor, C. (1972). On calculating with b-splines. *Journal of Approximation theory 6*(1), 50–62.

De Boor, C., C. De Boor, E.-U. Mathématicien, C. De Boor, and C. De Boor (1978). *A practical guide to splines*, Volume 27. Springer-Verlag New York.

deLeeuw, J. (1992). Introduction to akaike (1973) information theory and an extension of the maximum likelihood principle. *Breakthroughs in Statistics: Foundations and Basic Theory*, 599–609.

Eilers, P. H., I. D. Currie, and M. Durbán (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics & Data Analysis 50*(1), 61–76.

Eilers, P. H. and B. D. Marx (1992). Generalized linear models with p-splines. In *Advances in GLIM and Statistical Modelling*, pp. 72–77. Springer.

Eilers, P. H. and B. D. Marx (1996). Flexible smoothing with b-splines and penalties. *Statistical science*, 89–102.

Eilers, P. H. and B. D. Marx (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics 2*(6), 637–653.

Eilers, P. H., B. D. Marx, and M. Durbán (2015). Twenty years of p-splines. *SORT: statistics and operations research transactions 39*(2), 0149–186.

Fellner, W. H. (1986). Robust estimation of variance components. *Technometrics 28*(1), 51–60.

Fellner, W. H. (1987). Sparse matrices, and the estimation of variance components by likelihood methods. *Communications in Statistics-Simulation and Computation 16*(2), 439–463.

Haas, L. W. (1977). The effect of the spring-neap tidal cycle on the vertical salinity structure of the james, york and rappahannock rivers, virginia, usa. *Estuarine and Coastal Marine Science 5*(4), 485–496.

Heckman, N. E. and J. O. Ramsay (2000). Penalized regression with model-based penalties. *Canadian Journal of Statistics 28*(2), 241–258.

Horváth, L. and G. Rice (2014). Extensions of some classical methods in change point analysis. *Test 23*(2), 219–255.

Hurvich, C. M. and C.-L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika 76*(2), 297–307.

Imoto, S. and S. Konishi (2003). Selection of smoothing parameters inb-spline nonparametric regression models using information criteria. *Annals of the Institute of Statistical Mathematics 55*(4), 671–687.

Kielmas, M. (2019). How Does Salinity Affect the Solubility of Oxygen in Water? https://sciencing.com/info-10024026-salinity-affect-solubility-oxygen-water.html. [Online; accessed 9-March-2019].

Killick, R., P. Fearnhead, and I. A. Eckley (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association 107*(500), 1590–1598.

Lee, D.-J. and M. Durbán (2011). P-spline anova-type interaction models for spatio-temporal smoothing. *Statistical modelling 11*(1), 49–69.

Lee, T. C. (2003). Smoothing parameter selection for smoothing splines: a simulation study. *Computational statistics & Data analysis 42*(1-2), 139–148.

Mallows, C. L. (1973). Some comments on c p. *Technometrics 15*(4), 661–675.

McMullan, A., A. Bowman, and E. Scott (2003). Non-linear and nonparametric modelling of seasonal environmental data. *Computational Statistics 18*(2), 167–183.

Müller, H.-G., U. STADTMÜLLER, and T. SCHMITT (1987). Bandwidth choice and confidence intervals for derivatives of noisy data. *Biometrika 74*(4), 743–749.

Pan, W. (1999). Bootstrapping likelihood for model selection with small samples. *Journal of Computational and Graphical Statistics 8*(4), 687–698.

Ringler, N. H. and J. D. Hall (1975). Effects of logging on water temperature, and dissolved oxygen in spawning beds. *Transactions of the American Fisheries Society 104*(1), 111–121.

Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of computational and graphical statistics 11*(4), 735–757.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika 78*(4), 719–727.

Schoenberg, I. J. (1946). Contributions to the problem of approximation of equidistant data by analytic functions. part b. on the problem of osculatory interpolation. a second class of analytic approximation formulae. *Quarterly of Applied Mathematics 4*(2), 112–141.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics 6*(2), 461–464.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American statistical Association 88*(422), 486–494.

Simpkin, A. and J. Newell (2013). An additive penalty p-spline approach to derivative estimation. *Computational Statistics & Data Analysis 68*, 30–43.

Sugiura, N. (1978). Further analysts of the data by akaike's information criterion and the finite corrections: Further analysts of the data by akaike's. *Communications in Statistics-Theory and Methods 7*(1), 13–26.

Ugarte, M., A. Adin, and T. Goicoa (2017). One-dimensional, two-dimensional, and three dimensional b-splines to specify space–time interactions in bayesian disease mapping: Model fitting and model identifiability. *Spatial Statistics 22*, 451–468.

Wahba, G. et al. (1985). A comparison of gcv and gml for choosing the smoothing parameter in the generalized spline smoothing problem. *The Annals of Statistics 13*(4), 1378–1402.

Wang, T. and R. J. Samworth (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(1), 57–83.

Wood, S. (2005). GAMs with GCV smoothness estimation and GAMMs by REML/PQL. https://www.rdocumentation.org/packages/mgcv/versions/1.2-0. [Online; accessed 20-March-2019].

Wood, S. N. (2017). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC.

Wood, S. N., Y. Goude, and S. Shaw (2015). Generalized additive models for large data sets. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 64*(1), 139–155.

Xue, L. and L. Yang (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica 16*(4), 1423.