Statistical approaches for constructing runoff maps

Random forest, linear models and spatial models tested for predictive performance of ungauged catchments in Norway

Ida Jahren Herud

Supervisor: Ingelin Steinsland

Department of Mathematics University of Bergen

Master of Science in Statistics (Data Analysis) STAT399

June 3, 2019

Summary

In this thesis we study statistical approaches to tackle predictions of ungauged catchments in Norway. In collaboration with Norwegian Water Resources and Energy Directorate (NVE) we use observations of runoff, catchment characteristics (e.g. elevation and land use) and observations of precipitation for constructing runoff models suitable for runoff maps. The challenge of constructing suitable models for ungauged catchments is due to the lack of observations, and in the field of hydrology this problem is known as *the problem of ungauged basins* (Blöschl et al., 2013). It is common to either use a deterministic hydrological model or a suitable model for transferring observations from a gauged catchment to an ungauged catchment. Our statistical approach for modeling median annual runoff has been done in a three-step procedure where our main focus has been on the models predictive performance also including the uncertainty quantification. First, we did an exploratory analysis of observed median annual runoff and catchment characteristics. As a second step we fitted two initial model classes (linear regression models and random forests models) where we observed how the different explanatory variables/features influenced our predictions. With the main learning's of our second step we built four different spatial models within the Bayesian framework. From the main learning's we found that spatial dependency have a large effect on predictive performance, and that gradient basin was the only catchment characteristics that influenced the models. The model with the best predictive performance was a Bayesian hierarchical model of three levels where gradient basin was included in addition to a Gaussian random field (GRF) and precipitation with a spatially varying coefficient. All models have been carefully evaluated through leave-one-out cross-validation (LOOCV), where each model have been evaluated in terms of predictive performance with the two evaluation metrics; root mean square error (RMSE) and continuous ranked probability score (CRPS). While RMSE describes the difference between observed and predicted runoff, we account for the whole posterior predictive distribution with CPRS, and is thus useful for quantifying the uncertainty of our predictions.

Preface

This thesis is the final work of my master's degree in statistics with a specialization in data analysis at the University of Bergen (UiB). The work was carried out through fall 2018 and spring 2019.

I would like to thank my supervisor Ingelin Steinsland, for shearing her knowledge and always finding time to answer all my questions. I would also like to thank Koldbjørn Engeland at NVE for helpfully answering all my questions regarding the dataset.

Finally, I want to thank my boyfriend Tarald for the support thought this year, without you it would have been a lot less pleasant. I also want to thank my aunt Ellen and my mom Tone for their patient help through the last few months of my work. My team at Bouvet also needs a great thank for their support, the support have meant a lot to me and I am looking forward to join the team full time this fall. I am also grateful for the support of my family and friends, their encouragement and cheering have motivated me to keep up the hard work. I would also like to thank my self, I am very proud of what I have accomplished, and everything that I have learned while writing this thesis.

Ida Jahren Herud
Bergen, May 2019

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis is done in collaboration with the Norwegian Water Resources and Energy Directorate (NVE). NVE produces runoff maps describing mean annual runoff within Norway for 30 years periods (see Beldring et al. (2002)). Runoff maps describes the amount of water that flows trough a specific area within some time period. Runoff maps are important tool frequently used in the fields of hydro power, water supply, agriculture and engineering projects. In fact, 99 % of all power production in Norway comes from hydro power (Statkraft, 2016). Accurate predictions of runoff with uncertainty estimates can contribute to improvement of these runoff maps, and thus also planning of hydropower production.

Motivated by improving the predictive performance of models for runoff, we explore and develop statistical models for median annual runoff based on predictive performance. We demonstrate how observations of runoff in neighbouring catchments, catchment characteristics, and precipitation can be used for predictions of runoff in areas where observations of runoff and/or precipitation does not exist.

Our statistical models for predicting runoff are in hydrology known as models of regionalization. For our models we use catchment characteristics, observations of runoff and observations of precipitation, all provided by NVE. The observations of runoff comes from 266 catchments. For all catchments we require at least 10 years of daily observations between September 1st 1986 to August 31st 2017 to be included in the dataset. A hydrological year is from September 1st until August 31st the following year, such that storage effects from snow does not have to be considered. Each catchment has corresponding catchment characteristics describing catchments attributes e.g. elevation and land use. Observations of precipitation is also corresponding to each catchment, with the same requirements as for runoff. The precipitation data is referred to as SeNorge, and it is published at *http://www.senorge.no/*, and produced by The Norwegian Meteorological Institute (MET). For a thorough description of the SeNorge data we refer to Lussana et al. (2018). When we develop our statistical models we have considered median annual runoff and median annual precipitation, as the median is less affected by outliers than the mean.

A great challenge in the field of hydrology is known as *the problem of ungauged basins* (Blöschl et al., 2013). For predicting ungauged catchments (basins) it is common to either

use a hydrologic approach or a regionalization approach. A hydrologic model is often a deterministic model that uses precipitation and temperature as exploratory variables for estimating runoff. It is also common that hydrologic models are calibrated/optimized with local observations of e.g. runoff. A model of regionalization describes how one can transfer information from an observed catchment to an ungauged catchment. Regionalization approaches are defined by He et al. (2011) as either distance based approaches (spatial proximity and physical similarity) or regression based approaches.

Our approach is divided into a three-step procedure where we (1) perform an exploratory analysis of our catchment characteristics, observations of runoff and observations of precipitation. (2) we use multiple linear regression and random forest to investigate the relationships between runoff and catchment characteristics and also to investigate the effect of having observations of neighbouring catchments and precipitation. (3) we build spatial models based on our main learning's of (1) and (2). Note that for our models we refer to runoff as the dependent variable while catchment characteristics and observation of precipitation is referenced as independent variables and explanatory variables within the linear models and the spatial models. For random forest models catchment characteristics and observation of precipitation is referenced as features.

With multiple linear regression we assume that there is a linear relationship between runoff and explanatory variables that can be modeled with some random errors (Fahrmeir et al., 2013). Multiple linear regression is a common method used for predicting ungauged catchments (Parajka et al. (2005)). Most studies comparing (multiple) linear regression with other regionalization approaches find that multiple linear regression performs worse than other approaches (see e.g. Yang et al. (2018), Parajka et al. (2005) and Parajka et al. (2013)). As previous work indicates that multiple linear regression do not model runoff well, we use it as a tool for investigation of how well the different explanatory variables are for predicting runoff.

Our multiple linear regression models assumes a linear relationship between runoff and explanatory variables. With random forest we are able to explore non-linear relationships between runoff and features (explanatory variables) and also interaction between the different features. Random forest was first introduced by Breiman (2001), and segments the predictor space into a number of simple regions. Random forest are in most cases not competitive with the best supervised learning approaches (James et al., 2013). On the other hand random forest regression are easy to use and we can visualise how the features influence our prediction of runoff, it is thus used as a additional tool for exploring our features.

To our knowledge there are few studies using random forest for modelling runoff. Li et al. (2016) compare random forest with a linear model, artificial neural network and support vector regression. Their area of study is the Poyang Lake in China, and by comparing the different models in terms of R^2 and mean square error (MSE), they find that the random forest gives the most reliable results when predicting daily water level. The only other study found so far is the master thesis, White (2015), which compares random forest with the Basin Characterization Model (BCM) and a linear multivariate regression model for predictions of unimpaired flow in ungauged basins in 69 California basins. When they compare the models with the R^2 and Nash-Sutcliffe efficiency score, they conclude that the BCM is the best and the linear multivariate regression model is the worst.

Due to large local variations of runoff within Norway, we assume that neither multiple linear regression nor random forest regression are assumed to be suited models for predicting runoff. Although the work of Li et al. (2016) and White (2015) indicates that random forest have more accurate predictions for runoff than multiple linear regression, it is better to have a model where spatial dependency is accounted for.

Runoff is part of the climate system, and it is first and foremost a result of precipitation and some evaporation. We know that precipitation in Norway is influenced by the Gulf Stream and the large differences in elevation, which gives large local variations in space. It is thus reasonable to assume that neighbouring catchments are more related than distant catchments. For the initial models we use the average of the five closest catchments as a proxy for the spatial dependency, and precipitation is also included to explore the effect on our models predictive performance. Such spatial dependencies can also be modeled in a more direct manner by a spatial effect, and this is why we use spatial models to build our final models. Spatial models are within the field of geostatistics and in the past been done by Kriging approaches (see eg. Sauquet et al. (2000) and Skøien et al. (2006)). Roksvåg et al. (2019) use a Bayesian geostatistical model for interpolating hydrological data in the Voss area, similar work was conducted for modeling precipitation in Ingebrigtsen et al. (2015).

Motivated by the work of Roksvåg et al. (2019), our spatial models are linear mixed models where we allow the random effect to vary spatially. For computational benefits we use Bayesian linear mixed models with an hierarchical structure, these models are known as latent Gaussian model (LGM). Where the LGM is a subclass of hierarchical models within the Bayesian framework. Our model consist of three levels, where the first level specifies the likelihood of the observation given some parameters and hyper parameters, second level describes the probability of the spatial process given some parameters. The third level is the model for the parameters specifying the prior distribution of the hyperparameters. To draw inference for such models we use the Integrated nested Laplace approximation (INLA) (Rue et al., 2009). Our spatial effects are modeled by a Gaussian random field (GRF). As GRFs are computationally expensive we use the Stochastic Partial Differential Equation (SPDE) proposed by Lindgren and Rue (2011). The SPDE allows us to express our GRF as a Gaussian Markov random field (GMRF). GMRFs reduce the computational cost and allows for fast inference (Rue and Held, 2005).

For our spatial models it is also possible to allow the coefficient of the explanatory variables to vary spatially. This was done in the work of Gelfand et al. (2003) where they present spatial models with spatially varying coefficients. Motivated by their work we will use spatially varying coefficients in our spatial models, such that the importance of the explanatory variables depends on the location of a catchment.

We explore the models ability to make accurate predictions of runoff. For this we use leave-one-out cross-validation (LOOCV), where we use observations for all other catchments to do predictions of runoff in the catchment left out. We measure the predictive performance by calculating the mean of the root mean square error (RMSE) and the mean of the continuous ranked probability score (CRPS). RMSE is only able to evaluate the posterior predictive mean, while CRPS is able to account for the whole posterior predictive distribution.

The aim of this thesis is to present geostatistical models for predicting runoff in un-

gauged basins suited for runoff maps. Our main focus is on the models predictive performance. We also evaluate how the exploratory variables of catchment characteristics and precipitation influence the predictive performance.

The outline of our thesis is as follows: In chapter 2 we introduce the data received from NVE, it also consists of a thorough exploratory analysis of the observations and catchment characteristics. In Chapter 3 we introduce the background and underlying theory of the methods applied in the thesis. In chapter 4 we build our models, and in Chapter 5 we first present the results of the initial models, present the catchment characteristics ability to model runoff and whether information about neighbouring catchments should be introduced into the models. Further in chapter 5 we also do a brief summary of the main learning's of the initial models. In chapter 6 we evaluate the predictive performance of our spatial models, and also explore how the models perform for different locations and levels of runoff. We also investigate posterior distribution the coefficients, SPDE parameters and GRFs. In chapter 7 we discuss the results of our studies.

Chapter 2

Data and exploratory analysis

In this chapter we first introduce the data from 266 catchments used in this thesis and we follow up with an exploratory analysis. The exploratory analysis has been an important part of our work because it has prepared us for the main goal of this thesis–doing accurate predictions of ungauged basins that minimizes the uncertainty. Exploratory analysis of our data is essential for creating good models. It lets us understand relationships within the data and can help us answer and form new questions about the data.

For the exploratory analysis the app *https://idajahrenherud.shinyapps.io/shiny_app/* was built as a tool enabling visual exploration of the catchments used within our thesis. With this app we are able to explore a catchment location and characteristics. Throughout the thesis we refer to catchments by their field ID, which can be used as a key for locating catchments within the app.

The app has also been used as a tool for exploring spatial dependency between yearly observations of runoff. With this app it is possible to evaluate the correlation of the yearly runoff for one catchment with all other catchments. We also explored the spatial correlation between catchments through a variogram. As the variogram describes the degree of relationship between yearly runoff for all catchments it has been left out for readability of this chapter as we only explore and model the median annual runoff. The variogram with a corresponding semivariance plot can bee found in the appendix C.

2.1 Data

The collaboration with NVE has enabled us to work with daily observations of runoff from 699 catchments across Norway. The catchments we received observations from are displayed as points in **fig. 2.1**, here the points marked in blue belongs to the 266 catchments used in this thesis. The acquired runoff data (is supposed to) only represent unregulated catchments. NVE has done a thorough review and evaluated the quality and how well the raw data fits as a valid observation for the analysis. Only stations with data from after 1958 and with at least five years of data has been included. Catchments that have been

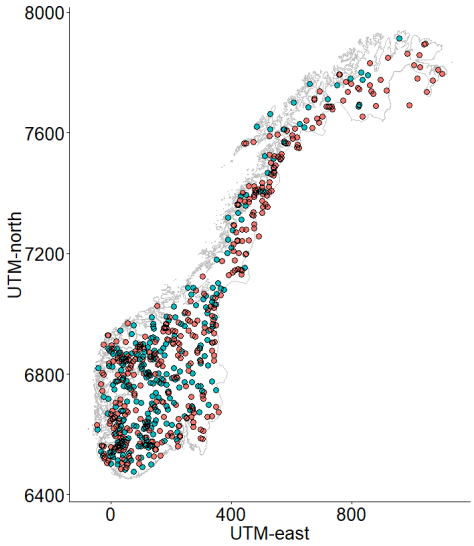


Figure 2.1: Map showing the centroid of the 699 catchments provided to us by NVE. Red points locates catchments that was left out of our analysis, and blue points locates the 266 catchments that was used.

regulated in the past, but not anymore, are also included, as for the catchments that today are regulated, but not in the past.

An initial analysis of the data received from NVE was conducted, this analysis led us to the 266 catchments used for our thesis. The initial analysis is presented in the appendix 2. All data selected for this thesis was done in collaboration with NVE. If any suspect observations were found we had a discussion with NVE regarding the observation, in order to decide whether it should be removed or not.

The unit of the observed runoff data from NVE was received in the unit m^3/s. The unit of our work is in mm/year. To transform our data into the preferred unit, each daily observation was divided by the area of catchment and multiplied by 86.4, giving mm/day. We accumulated each year, which returns mm/yr. The time period of our thesis is from September 1st 1986 until August 31st 2017. In hydrological years this is 1986 to 2016.

The notation used for the 266 observations of runoff are $y_{i,t}$ where i denotes catchment and t denotes year. Our models are based on median yearly runoff observations which we denotes as $\tilde{y}_i$ for catchment i. Median yearly runoff is the median of all yearly runoff observation $y_{i,t}$ for each catchment i. We refer to runoff as median annual runoff $\tilde{y}_i$ from here on. We use the centroid of all catchments for point predictions, is denoted as s_i for catchment i.

From NVE we also received catchment characteristic, this is denoted as x_{ji} for characteristic j, catchment i. The catchment characteristics used for analysis are listed in **Tab. 2.1**.

We received the precipitation as daily observations for all catchments, and as for runoff we summed each year and found the median observation within our time period. Observa-

tions of median annual precipitation is denoted as p_i, and often referred to as precipitation. The precipitation data is a product of interpolation over a high resolution grid (1 km spacing) that is updated daily.

2.2 Software

The data used in the analysis is of various data formats, the runoff observations is gathered from multiple text files, data with catchment information is stored in a large Excel file, and map data for catchments and stations are stored as shape files.

R by R Core Team (2013) has been used as programming language. It provides a variety of statistical packages that have been useful when getting all the data in the right format, transforming, plotting, mapping, fitting models and evaluating the models.

For merging all the text files, the package *dplyr* has been of great help (Wickham et al., 2017). This package has been used to restructure data and also transforming data.

The catchment characteristics that was received in an Excel file which was imported via a simple base function in *R*. The rows that contained catchments that did not contain sufficient information was removed. Further a data frame with the median discharge was appended, as the observation data used for further analysis.

The spatial data where with help of the *sp* package (E. J and Bivand, 2005; Bivand et al., 2013) and the *rgdal* package (Bivand et al., 2016) packages imported and transformed to the preferred coordinate system. For mapping with the *leaflet* package (Cheng and Xie, 2016) a long-lat projection was used. Calculations of distance was done with a UTM 33 projection of our coordinates, this returns distance in meters. Enabling us to view catchment characteristics and observations in a map, a data frame was merged with the *SpatialPolygonsDataframe* where the field ID identifies what observations belongs to which polygon.

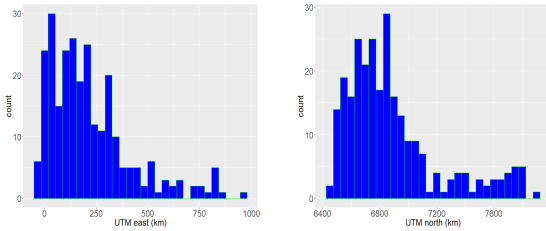
2.3 Exploratory analysis of Catchment characteristics

We now present an exploratory analysis of the catchment characteristics. The catchment characteristics are presented in **table. 2.1**. We first present the distribution of characteristics within the catchments. For this we subdivide the different catchment characteristics into sections that contains characteristics describing similar attributes of a catchment. Further we investigate if there are a linear association between the individual catchment characteristics.

2.3.1 Easting and northing

Easting and northing refers to the UTM east and -north coordinates of the centroid of the catchments. UTM is short for Universal Transverse Mercator, and is a two dimensional Cartesian coordinate system. UTM east is the distance from the central meridian in the relevant UTM zone and UTM north is the distance from equator, we use UTM zone 33.

The histograms describing the distribution of UTM east and UTM north coordinates are displayed in **fig. 2.2**. Most UTM east coordinates are located at small values, as seen

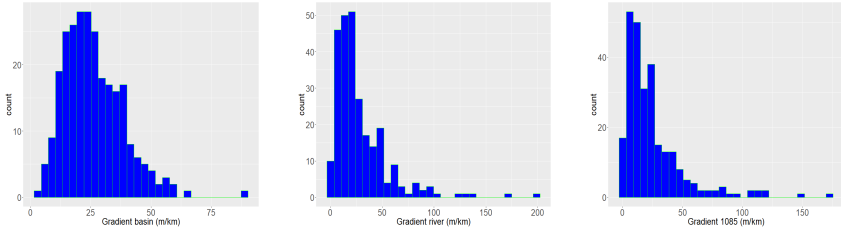


(a) Histogram of UTM east. **(b)** Histogram of UTM north.

Figure 2.2: Histogram of the UTM coordinates.

in the histogram in **fig. 2.2a**. The histogram illustrating UTM north coordinates (**fig. 2.2b**) shows that most catchments are also located around 6000 km north. If we have a look at **fig. 2.1** it can be seen that most catchments are located in the southern parts of Norway while less catchments are located in the northern parts. Northern Norway has a rotation towards the east, which is why most UTM east coordinates are small.

2.3.2 Gradients



(a) Gradient basin (m/km). **(b)** Gradient 1085 (m/km). **(c)** Gradient river (m/km).

Figure 2.3: Histogram of gradients used for this thesis. Notice that the range of the gradients differ.

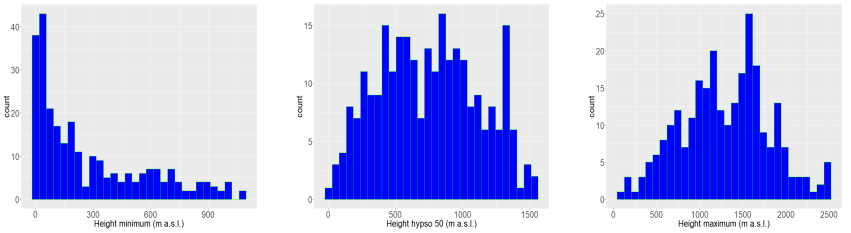
Gradient river and gradient 1085 are both measures of how the difference is in height of the main river within a catchment. Gradient river is the total difference in height along the main river divided by the length of the river. Gradient 1085 is similar but it is exclusive the 10% lowest and 15% highest parts of the river. Gradient basin is the total difference in height within the catchment. The unit of the three measures are in m/km.

Gradient basin (**Fig. 2.3a**) has a mean of 26.15 m/km, with one outlier that has a gradient of 89.6. This belongs to *Lundberg* (field ID 2082), and is located north, in the mountainous area of *Bardufoss*. Gradient river in **fig. 2.3c** has a mean gradient of 28 m/km, but reviles some steeper gradients, with a maximum of 200 m/km, which belongs to *Nigardsbreavatnet* (field ID 1339), which is located in the mountains of the north-west part of Norway. Gradient 1085 in **fig. 2.3b** is similar to gradient river with a mean of 28 m/km, here maximum is at 173 m/km. The larges gradient river belongs to *Engabrevatn* (field ID 1893), and is at the cost in the northern parts of Norway. In **Fig. B.1** maps of

the different gradients can be seen. We see how the gradients are higher towards west and north, as these are mountainous areas.

All measures of gradient reviles that most catchments contains some elevation change. This is as expected since Norway is dominated by mountains and most catchments contains some mountainous areas, but there are more mountains towards the east and along the coast. As gradient river and gradient basin describe the total difference in elevation of the main river they revile much larger changes in elevation, than gradient basin that account for the total difference in elevation for the whole catchment, which make gradient basin more symmetrically distributed than gradient river and gradient basin.

2.3.3 Elevation



(a) Height minimum. **(b)** Height hypso 50. **(c)** Height maximum.

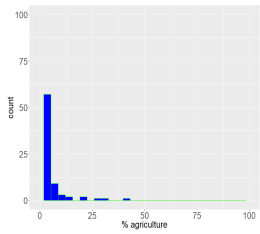
Figure 2.4: Histogram of elevation (m a.s.l).

Our catchment characteristics height minimum, height hypso 50 and height maximum are all measures of how far above sea level the catchment is located in meters. Height minimum is the lowest elevation point within the catchment, height hypso 50 is the 50% percentile for the hypsometric curve for our catchment. The hypsometric curve is described by Strahler (1952) as a curve that relates horizontal cross-sectional area of a catchment to the relative elevation above the catchments lowest point. Height maximum is the highest elevation point within our catchment.

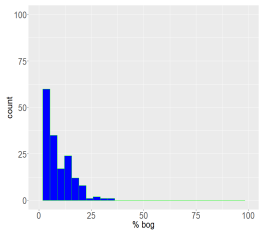
Fig. 2.4a shows how the minimum height for most catchments centres around 50 meters above sea level (m a.s.l.) but we also have a long upper tail ranging from 100 m a.s.l. up to 1077 m a.s.l. Height hypso 50 is displayed in **Fig. 2.4b** showing a more symmetrical distribution than height minimum. Height hypso 50 has a mean of 752 m a.s.l. Height maximum in **Fig. 2.4c** also shows a symmetric distribution of observations and has a mean of 2463 m a.s.l. The large values of elevation are found in the interior of southern Norway as we can see from the maps in **Fig. B.2**. **Fig. B.2a** shows that catchments along the eastern border of Norway have smallest minimum height, **Fig. B.2c** also shows how some of the catchments along the coast has large maximum heights.

The heights illustrates what we had expected, as the largest elevations are found in the interior in the southern parts of Norway. For the catchments in the interior, both maximum and minimum elevation are larger than most catchments. And for catchments along the coast and along the border of Sweden we have the minimum elevation at approximately sea level and do not have as large maximum height as seen for interior catchments.

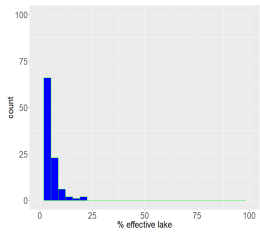
2.3.4 Land characteristic ratios



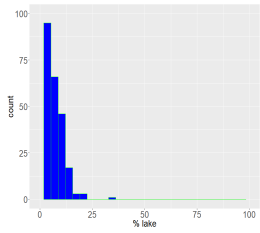
(a) Percentage agriculture.



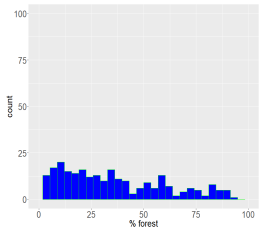
(b) Percentage bog.



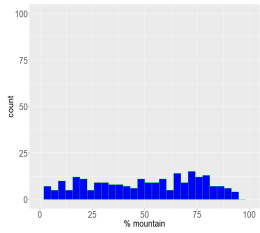
(c) Percentage effective lake.



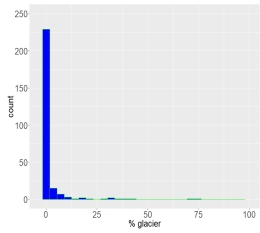
(d) Percentage lake.



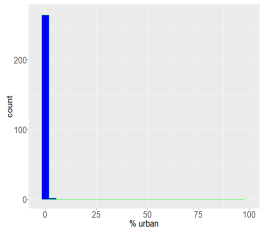
(e) Percentage forest.



(f) Percentage mountain.



(g) Percentage glacier.



(h) Percentage urban.

Figure 2.5: Histogram of land characteristic ratios.

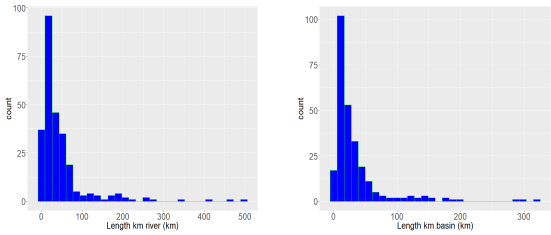
For ratios of land characteristics we can determine what types of characteristics that dominates within each catchment. The different ratios can be seen in **fig. 2.5**. The different land characteristics are agriculture, bog, effective lake, lake, forest, mountain, glacier and urban.

fig. 2.5a shows that percentage of agriculture is small as most catchments contains no agriculture and only six catchments contains more than 20% agriculture, with a maximum of 42%. Percentage of bog in **fig. 2.5b** shows that most catchments contains some bog, with an average of 5%. Percentage of effective lake in **fig. 2.5c** has an mean of 2.17%, and with a small upper tail. **Fig. 2.5d** show percentage of lakes, with a mean of 6.17% with an outlier at 35% named *Storvatn* which is located in the south western parts of Norway. Percentage forest in **fig. 2.5e** shows a large range of different observations, with a mean of 34%, a maximum of 100% and minimum of 0%, showing that most of the catchments contains some forest. **Fig. 2.5f** also shows that most catchments contains some mountains. **Fig. 2.5g** shows how a few catchments has some glaciers, but most do not, we see two outliers which is *Nigardsbrevatn* (field ID 1339), with 73% glacier and *Engabrevatn* (filed

ID 1893), with 72% glacier. In **Fig. 2.5h** catchments with urban areas have an average of 0.08 %, showing that this does not dominate any catchments.

In **fig.B.3** all the different ratios of land characteristics are displayed in different maps. Here we see how most of the characteristics only account for a small percentage of the catchments attributes. The only dominating characteristics are percentage forest in **fig. B.3e** and percentage mountain in **fig. B.3f**. As the altitude increase percentage of forest decrease and percentage of mountains increase, such that the the compliment each other. As expected we can also see how areas in south eastern parts of Norway has the most forest, and that western and coastal parts contains the majority of mountains.

2.3.5 Rivers and basins



(a) Length river (km). **(b)** Length basin (km).

Figure 2.6: Histogram showing lengths of rivers and basins.

Length river is the total length of all the rivers within a catchment, and length basin is the total length of the catchment.

In **Fig. 2.6b** length of the basins is centred around 33 km, while some catchments are much longer. For example *Rånåsfoss* (field ID 248), which is the longest catchment in our data has a length of 321 km, and is located in the south eastern part of Norway. The rivers length in **fig. 2.6a** has a similar shape as the length of the basins, with most observations centred around 45 km. The longest river also belongs to *Rånåsfoss*. In **Fig. B.4** *Rånåsfoss* is easy to spot, and we can see how the longest rivers and basins are found in the south eastern part of Norway.

2.3.6 Dependency between catchment characteristics

In the following section we explore the dependency between our catchment characteristics. If catchment characteristics are highly correlated and used within the same model it could affect the models predictive performance. The correlation matrix in **Fig. 2.7** represents the correlation coefficient which indicates whether two catchment characteristics are linearly related. As several catchment characteristics describes natural phenomenons, it is reasonable to assume that some strong correlations are observed.

Fig. 2.7, shows that several catchment characteristics are highly correlated. The figure shows for example how all the characteristics describing elevation (height_min, height_hypso_50 etc.) have high positive correlations with each other. We also see that

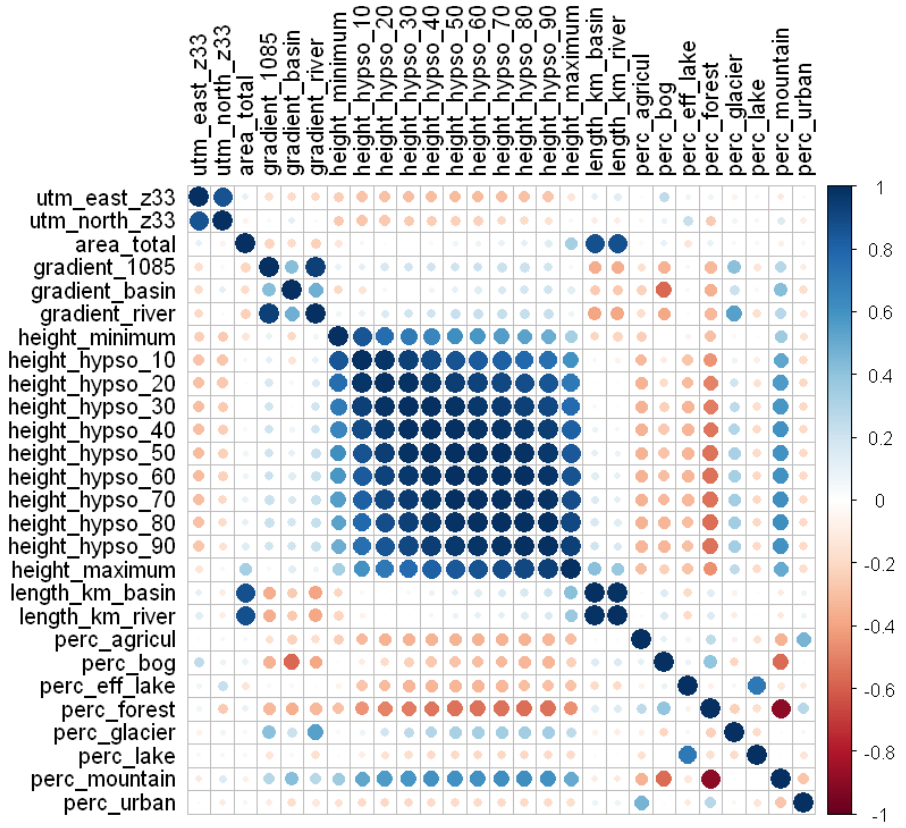


Figure 2.7: A plot showing correlation between catchment characteristics. The blue colour represents a positive correlation and the red colour represents a negative correlation. The correlation values are calculated with correlation coefficient, which measures the linear dependency between the catchment characteristics.

the percentage of forest within a catchment has strong negative correlation with elevation. The strong negative correlation is reasonable since we know that increasing elevation leads to a decrease in forest. This is also reflected in the high negative correlation between the percentage of forests and the percentage of mountains within the catchments.

The correlation matrix (**Fig. 2.7**) shows how most catchment characteristics have some degree of linear association. And as expected, the characteristics describing similar physical attributes of a catchment have a positive linear association.

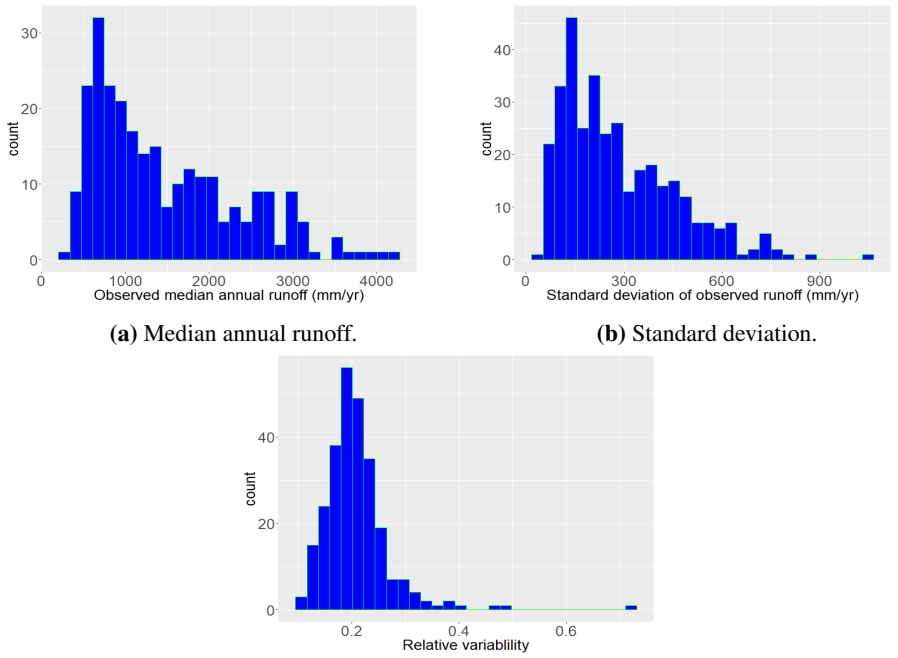
Name	Catchment feature	Unit
utm_east_z33	east UTM 33 coordinates	m
utm_north_z33	north UTM 33 coordinates	m
area_total	total catchment area	km^2
gradient_1085	river gradient exclusive 10% and 85% elevation of river	m/km
gradient_basin	basin gradient	m/km
gradient_river	river gradient	m/km
height_minimum	minimum height	m a.s.l.
height_hypso_10	10% percentile for hypsographic curve	m a.s.l.
height_hypso_20	20% percentile for hypsographic curve	m a.s.l.
height_hypso_30	30% percentile for hypsographic curve	m a.s.l.
height_hypso_40	40% percentile for hypsographic curve	m a.s.l.
height_hypso_50	50% percentile for hypsographic curve	m a.s.l.
height_hypso_60	60% percentile for hypsographic curve	m a.s.l.
height_hypso_70	70% percentile for hypsographic curve	m a.s.l.
height_hypso_80	80% percentile for hypsographic curve	m a.s.l.
height_hypso_90	90% percentile for hypsographic curve	m a.s.l.
height_maximum	maximum height	m a.s.l.
length_km_basin	length of river	km
length_km_river	length of basin	km
perc_agricul	percentage agriculture	%
perc_bog	percentage bog	%
perc_eff_lake	percentage effective lake	%
perc_forest	percentage forest	%
perc_glacier	percentage glacier	%
perc_lake	percentage lake	%
perc_mountain	percentage mountain	%
perc_urban	percentage urban	%

Table 2.1: Table of all catchment characteristics received from NVE with a small description of what they are and what unit they have.

2.4 Exploratory analysis of observed runoff

Median annual runoff is plotted in a histogram in **fig. 2.8a**. The histogram shows how most observations of runoff are approximately 1000 mm/yr, while the long upper tail shows that many catchments have much larger observations of runoff. We observe that there are two observations that are larger than 4000 mm/yr, these are *Straumstad* (filed ID 1434) and *Flostrand* (field ID 1888), they are both located in wet coastal areas where it is reasonable to have large observations.

The standard deviation plotted in **fig. 2.8b** shows a similar distribution as median annual runoff with left skewed values and a long upper tail. The mean standard deviation is 284 mm/yr with a minimum of 50 mm/yr and a maximum of 1061 mm/yr. The catch-



(a) Median annual runoff.

(b) Standard deviation.

(c) Relative variability.

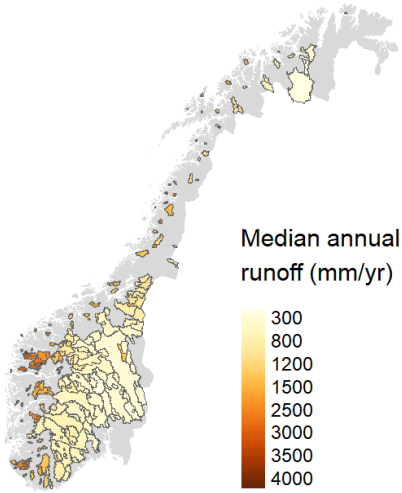
Figure 2.8: Histogram of observed median annual runoff (mm/yr), standard deviation annual runoff (mm/yr) and relative variability for the 266 catchment used in this thesis in the hydrological period 19872017.

ment with the larges standard deviation is *Engabrevatn* (field ID 1893). This catchment is located at *Svartisen* which is a large glacier. For glaciers it is natural to assume large deviations, as runoff in areas with glaciers depend on how much the glacier melts within a hydrological year.

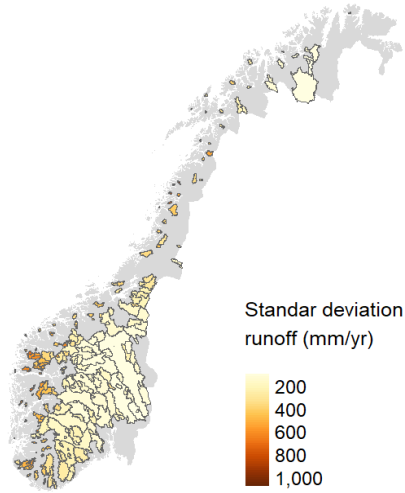
The relative variability plotted in **fig. 2.8c** is a measure of how much the standard deviation deviates from the observed median annual runoff. Most observations have a small relative variability with one observation that is far greater than all other observations. This belongs to *Engabrevatn* (feild ID 1893), which contains large amounts of glaciers. Observations of large relative variability is common for catchments located in areas with glaciers. One catchment not containing any glaciers, namely *Vismunda* (field ID 270) has a large relative variability. This is not expected as most catchments in this area have a low median annual runoff and a small standard deviation.

We can further explore how our observations of runoff are distributed by visualizing them in a map. This is done in **fig. 2.9**. The median annual runoff (**fig. 2.9a**) shows how the highest runoffs are to be found along the coast of Norway. The interior has a low runoff compared to the coastal areas. Towards north the largest observations are along the coast while smaller observations of runoff are found towards the eastern border.

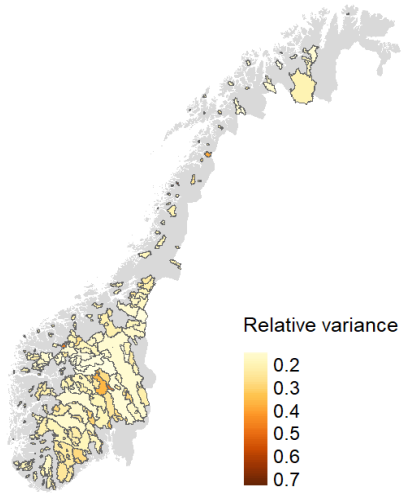
The standard deviation mapped in **fig. 2.9b** shows similar results as median annual



(a) Median annual runoff.



(b) Standard deviation.



(c) Relative variability.

Figure 2.9: Maps of median annual runoff (mm/yr), standard deviation annual runoff (mm/yr) and relative variability for the 266 catchment used in this thesis in the hydrological period 19872017.

runoff. Catchments with large standard deviation are located along the cost and towards the north. The relative variability mapped in **fig. 2.9c** shows that most catchments have similar relative variability, but some catchments in the interior of southern Norway have higher relative variability than most catchments.

The large observations of runoff along the coast are as expected, caused by the steep

topography leading to increased precipitation. The coastal areas are wet areas compared to the catchments in the interior, and as they are more wet they also have a larger standard deviation. The large observations of runoff does not indicate a larger relative variability, the catchments with larger relative variability are located in the interior and could be caused by the watershed separating western and eastern parts of southern Norway.

From the maps in **fig. 2.9** we also notice that the area of our catchments vary greatly. Within the south-east parts of Norway we have some large catchments ranging up to 38 440 km^2. For the coastal areas in the western and northern parts of Norway we have much smaller catchments, and our smallest catchment has a area of only 0.44 km^2.

2.5 Exploration of dependency between runoff and field characteristics

Now we explore how the different catchment characteristics are able to explain median annual runoff. We first investigate whether there are a linear association between median annual runoff and the individual catchment characteristics. Secondly we visually explore the relationship between individual catchment characteristics and runoff.

2.5.1 Linear dependency between catchment characteristics and runoff

The linear relationship between a runoff (dependent variable) y_i and a catchment characteristic (independent variable) x_i can be seen as

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, ..., n \tag{2.1}$$

where β_0 is the intercept, β_1 is the unknown regression coefficient of x_i and ϵ_i is assumed to be independent and identically distributed (i.i.d.) with mean 0 ($\mu = 0$) and variance σ^2. From this linear relationship we are able to investigate how our independent variable x_i is able to describe the dependent variable y_i. This is done with the p-value and the correlation coefficient. We now present an overview of the p-value and the correlation coefficient and how the two are related. The presentation is based on Devore and Berk (2012).

From the linear relationship described in **eq. 2.1**, we are able test the probability of there being a correlation between our dependent variable y_i and the independent variable x_i, known as the p-value. For estimating the p-value we assume that there is no correlation between the two variables y_i and x_i. Further we assume that the a test statistic (t-value) has a t distribution with n - 2 degrees of freedom. With this we calculate the p-value as the probability of the t-value being grater than zero, e.g. no correlation. If the calculated p-value is smaller than some level α we say that our p-value is significant, which means that there is a small probability of there not being any correlation between the two variables y_i and x_i, but it does not indicate how the relationship is.

With the correlation coefficient we are able to determine the direction of correlation. By direction we refer to a increasing or decreasing relationship between the dependent variable y_i and the independent variable x_i. For estimating the correlation coefficient we use the covariance, which describes the dependency between our two variables y_i and x_i,

e.g. $\text{Cov}(x_i, y_i) = E[(x_i - \mu_x)(y_i - \mu_y)]$. With the covariance $\text{Cov}(x_i, y_i)$ we obtain the correlation coefficient for our two variables y_i and x_i as follows,

$$r = \frac{\text{Cov}(x_i, y_i)}{\mu_x \mu_y} \tag{2.2}$$

where r is the correlation coefficient, and ranges between -1 and 1. Thus a correlation coefficient of ± 1 would indicate a perfectly positive/negative linear relationship between the dependent variable y_i and the independent variable x_i, while 0 means that there is no correlation.

For describing the relationship between our p-value and correlation coefficient we use the properties of the t-value, which are based on the assumption that the variables y_i and x_i are normally distributed. With the assumption of normally distributed variables, the t-value can be calculated as

$$\text{t-value} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}. \tag{2.3}$$

where n is the number of observations y_i for $i = , ..., n$. **Eq. 2.3** reflects the relationship between the p-value and the correlation coefficient. Telling us that a large absolute correlation coefficient returns a large t-value which again returns a small p-value.

With a significant level of $\alpha = 0.001$ we see from **tab. 2.2** that most catchment characteristics have a p-value smaller than our significance level α, while only seven catchment characteristics have a correlation coefficient more positive or more negative than ± 0.35. Further we observe that gradient basin have a correlation coefficient of 0.67, and it is thus the catchment characteristic that seem to have most linear association with runoff. We also notice that the two other characteristics describing the gradient within a catchment have a strong correlation coefficient, and as all three where found to have a strong linear association in the correlation matrix in **fig. 2.7** it is sufficient to only include one in our finale models.

From the results of the p-value and correlation coefficient listed in **tab. 2.2** we can assume that most catchment characteristics does not have a linear association with runoff.

2.5.2 Scatterplots

We explored the linear association in section 2.5.1. With the scatter plot presented in this section we visually explore the relationship between catchment characteristics and runoff. This allow us to investigate other relationships than the linear relationship. We note that if a relationship between a individual catchment characteristic and runoff are not visible from the scatter plot, it is still possible that there would be a relationship if we had some interaction between two or more characteristics. We fitted a simple linear regression line as presented in **eq. 2.1**, to each plot as it lets us associate the relationships between individual catchment characteristics and runoff to a linear relationship.

The UTM east and UTM north coordinates are plotted against runoff in **fig. 2.10**. For the UTM east coordinates plotted in **fig. 2.10a** there is a decrease in runoff with decreasing UTM coordinates, this is as expected, as we know that much of the catchments along the coast in southern Norway has large observations of runoff. For the UTM east coordinates

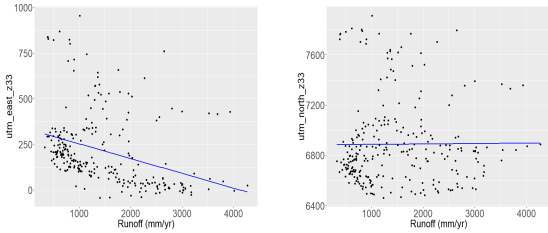
	catchment characteristic	correlation	p-value
1	**utm_east_z33**	**-0.3503**	0.000
2	utm_north_z33	-0.009	0.889
3	area_total	-0.2366	0.004
4	**gradient_1085**	**0.4298**	0.000
5	**gradient_basin**	**0.6760**	0.000
6	**gradient_river**	**0.5157**	0.000
7	height_minimum	-0.1087	0.077
8	height_hypso_50	-0.0059	0.923
9	height_maximum	-0.0482	0.433
10	length_km_basin	-0.3414	0.000
11	length_km_river	-0.3406	0.000
12	perc_agricul	-0.1827	0.015
13	**perc_bog**	**-0.4378**	0.000
14	perc_eff_lake	0.1897	0.003
15	**perc_forest**	**-0.4722**	0.000
16	perc_glacier	0.2434	0.000
17	perc_lake	0.1682	0.010
18	**perc_mountain**	**0.4223**	0.000
19	perc_urban	-0.1826	0.009

Table 2.2: A table with Person correlation coefficient calculated between catchments observed median annual runoff and characteristics, and the p-value from simple linear regression with one catchment characteristic. The catchment characteristics with a correlation larger than the absolute value of 0.35 correlation coefficient is written in bold.

in **fig. 2.10a** we also notice how some observations follow the same decreasing trend, but for larger values of UTM east coordinates, this belongs to the catchments in the northern parts of Norway as seen in the map in **fig. 2.9a** showing runoff. For the UTM north coordinates plotted against runoff in **fig. 2.10b** it is difficult to see any relationship as most, both large and small UTM north coordinates, contain large and small observations of runoff.

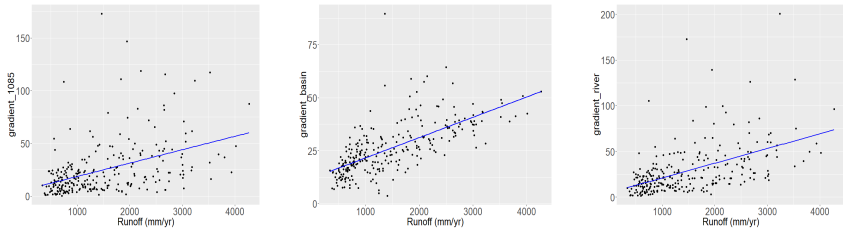
For the gradients plotted against runoff in **fig. 2.11** there seem to be some linear relationship with observed runoff. Gradient 1085 plotted against runoff in **fig. 2.11a** shows that there is a positive trend with observed runoff, but the deviation between some observations and the fitted line are large. Gradient basin plotted against runoff in **fig. 2.11b** shows a stronger relationship with median annual runoff compared to gradient 1085, were the deviation between observations is not as large as for gradient 1085. For gradient basin in **fig. 2.11b** we also notice that there is one outlier that deviates from the other observations. This belongs to *Lundberg* (field ID 2082), and is the catchment with the largest observed gradient basin (see section 2.3.2). Gradient river plotted against runoff in **fig. 2.11c**, show a similar scatter as seen for gradient 1085. As expected from the p-value and correlation seen in **tab. 2.2** there is a stronger linear association between gradient basin and runoff, than for gradient 1085 and gradient river.

The height observations plotted against runoff are seen in **fig. 2.12**, and neither show



(a) UTM east vs. runoff. **(b)** UTM north vs. runoff.

Figure 2.10: Scatter plot of runoff vs. UTM east and UTM north with a fitted linear regression line.



(a) Gradient 1085 vs. runoff. **(b)** Gradient basin vs. runoff. **(c)** Gradient river vs. runoff.

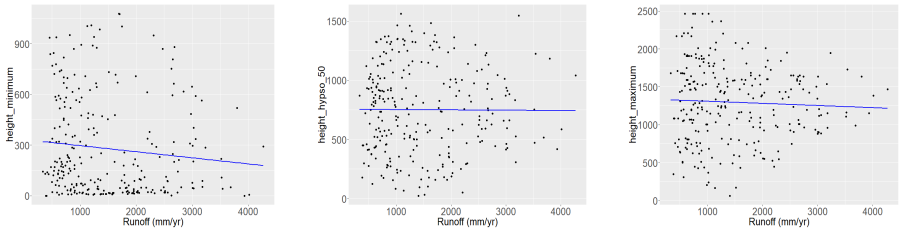
Figure 2.11: Scatter plot of runoff vs. gradients with a fitted linear regression line.

any relationship with runoff. As the p-value and correlation coefficient indicated in **tab. 2.2** it does not seem to be any linear association between the observations of height and runoff.

Moving on to the characteristics describing length we see from the plots in **Fig. 2.13** how there is a covariability between length km basin and length km river. From the plots (**Fig. 2.13**) we also notice that there are visibly not possible to determine a relationship between the lengths and runoff. And there is no linear association between the lengths of the river and basin and observed runoff, which we expected from the results of our exploration of linear dependency in section 2.5.1.

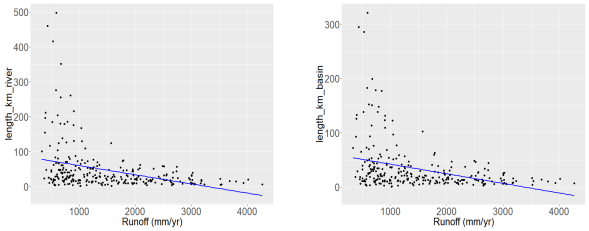
The catchment areas vary greatly. Therefore we also investigate whether there are some linear relationship between the area of our catchments and runoff. **Fig. 2.14** shows the area of the catchments plotted against runoff, and it does not seem to be any linear association here, as the fitted line does not describe our observations well. What we can see from **fig. 2.14** is that small catchments contains the largest observations of runoff.

Fig. 2.15 shows the different ratios of land characteristics plotted against runoff. Several land characteristics only account for zero to ten percent of a catchment attribute as we see in their histograms in **fig. 2.18b**. As the characteristics account for small portions of a catchment attribute they do not visually show any relationship with runoff which we can see in **fig. 2.15** were they are plotted against runoff. The only characteristics that dominates a catchment are forests and mountains. The percentage of forest is plotted against runoff in **fig. 2.15d**, the fitted line shows an increasing runoff with decreasing forests, but the deviation between observations is large and therefore it does not visually indicate



(a) Height minimum vs. runoff. **(b)** Height hypso 50 vs. runoff. **(c)** Height maximum vs. runoff.

Figure 2.12: Scatter plot of runoff vs. elevations with a fitted linear regression line.



(a) Length river vs. runoff. **(b)** Length basin vs. runoff.

Figure 2.13: Scatter plot of median observed annual runoff vs. lengths of rivers and basins with a fitted linear regression line.

a strong linear association. The same yields for percentage of mountain plotted against runoff in **fig. 2.15g**, where we have a positive increase, but large deviation between the observations. We notice from the plot of forest in **fig. 2.15d** and the plot of mountain in **fig. 2.15g** that the correlation between them found in the correlation matrix in **fig. 2.7** is reflected since they are opposite with a similar distribution of observations, this effect is known as covariability.

2.6 Average neighbour runoff

We assume that neighbouring catchments are more related than distant catchments due to some spatial correlation between our observations of runoff. With the assumption of spatial correlation we create a variable named avg_5, which allow us explore if it is reasonable to assume that the observations of runoff are spatially dependent (correlated). Avg_5 is used as an explanatory variable for the linear models and the random forest models. By comparing the predictive performance of the initial models with and without avg_5 we can explore spatial correlation between our observations of runoff.

We calculate avg_5 by calculating the distance from the centroid of one catchment to the centroid of all other catchments. The observations of median yearly runoff from the five catchments with smallest separation distance are then averaged and this returns avg_5_i for the $i = 1, ..., n$ observations of runoff.

We have illustrated avg_5 in **Fig. 2.16**, and from the map we observe the distribution

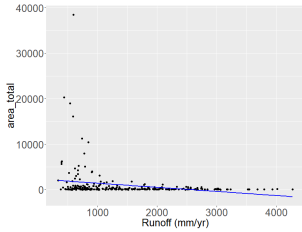


Figure 2.14: Scatter plots of runoff vs. area total with a fitted linear regression line.

avg_5 within Norway. We notice that the map of avg_5 is similar with the map illustrating the distribution of median annual runoff within Norway from **fig. 2.9a**. The histogram of avg_5 in **fig. 2.16b** illustrates how most observations are around 1000 mm/yr and we also have some observations that ranges up to 3500 mm/yr which is similar with our observations of median annual runoff in the histogram in **fig. 2.8**.

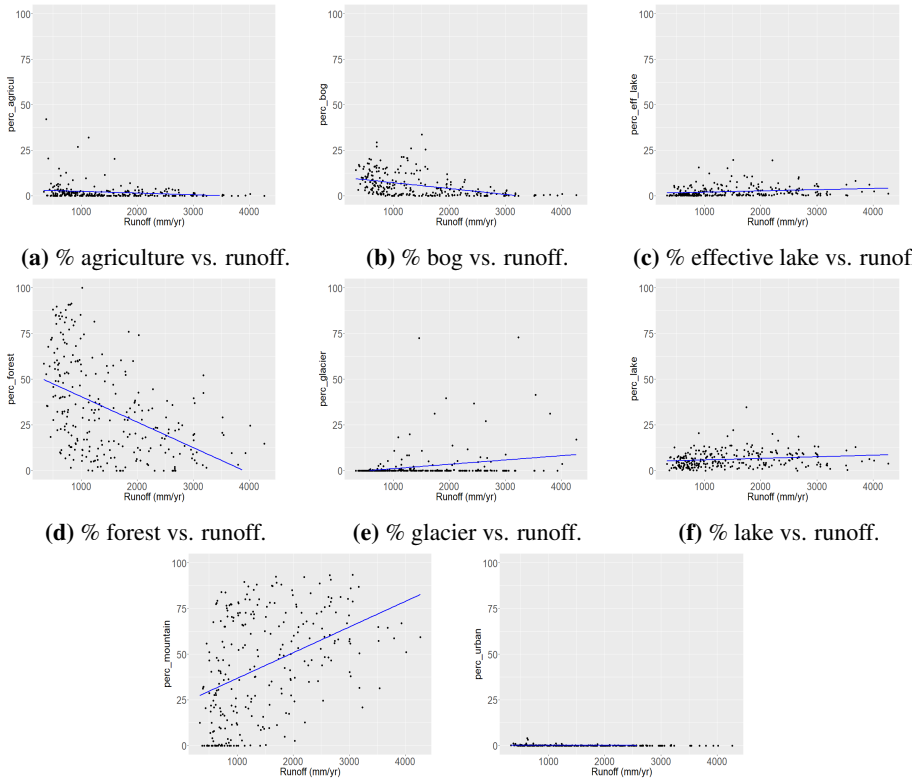
We fit a simple linear regression model as described in **eq. 2.1**, with runoff being the dependent variable and avg_5 as the independent variable. The simple linear regression model is plotted as a line in **fig. 2.17** where we see how there is a positive linear association between avg_5 and runoff. There is some deviation between the observations, but there is an increase in runoff with increasing avg_ 5. From the fitted model the resulting p-value is significant with a p-value of $2 * 10^{-16}$, and when calculating the correlation coefficient between median annual runoff and avg_5 we get a positive correlation coefficient of 0.83, which is much stronger than any correlation coefficient calculated between runoff and a individual catchment characteristic (see **tab 2.2**). The strong linear association between runoff and avg_5 indicates that it is reasonable to assume some spatial dependence between our observations of runoff.

2.7 Exploratory analysis of observed precipitation

With observations of precipitation we explore how the predictive performance change as it is introduced as an explanatory variable in our models for predicting runoff. We will now explore observations of precipitation. We know that precipitation is the driving force of runoff, and it is thus reasonable assume a strong relationship between the two.

The observed precipitation is plotted in a map in **fig. 2.18b** and shows that the largest observations of precipitation are located on the coast and in western parts of Norway. From the histogram in **fig. 2.18b** we observe that observed precipitation are centred around 1000 mm/yr and ranges up to approximately 4300 mm/yr. The minimum observed runoff is 518 mm/yr with a mean of 1321 mm/yr and has a maximum of 4340 mm/yr, which is similar with the observed values of runoff (see section 2.4).

If we fit the simple linear regression model from **eq. 2.1** with runoff as the dependent variable and runoff as the independent variable, we obtain the fitted line plotted in **fig. 2.19**. The corresponding p-value is $2*10^{-16}$ which is significant and the correlation coefficient is 0.88, which is even stronger than the correlation coefficient of avg_5 (see section 2.6). The significant p-value and the positive correlation coefficient indicates a linear association



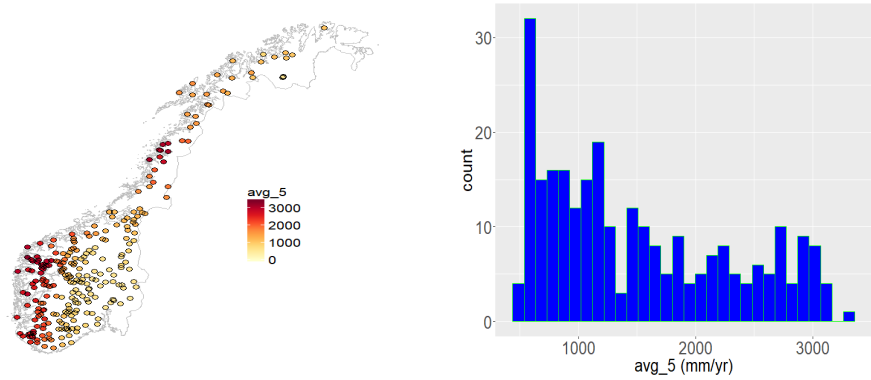
(a) % agriculture vs. runoff. **(b)** % bog vs. runoff. **(c)** % effective lake vs. runoff.

(d) % forest vs. runoff. **(e)** % glacier vs. runoff. **(f)** % lake vs. runoff.

(g) % mountain vs. runoff. **(h)** % urban vs. runoff.

Figure 2.15: Scatter plots of runoff vs. rations of land characteristics with a fitted linear regression line.

between runoff and precipitation, which is confirmed by the plot in **fig. 2.19** showing precipitation plotted against runoff. It is thus reasonable to assume precipitation and runoff are highly dependent.



(a) Map of avg_5. **(b)** Histogram of avg_5.

Figure 2.16: Map and histogram of avg_5, which is constructed form the average of the observed median annual runoff of the five closest neighbouring catchments.

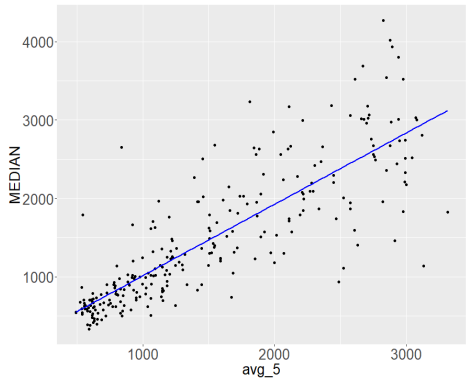
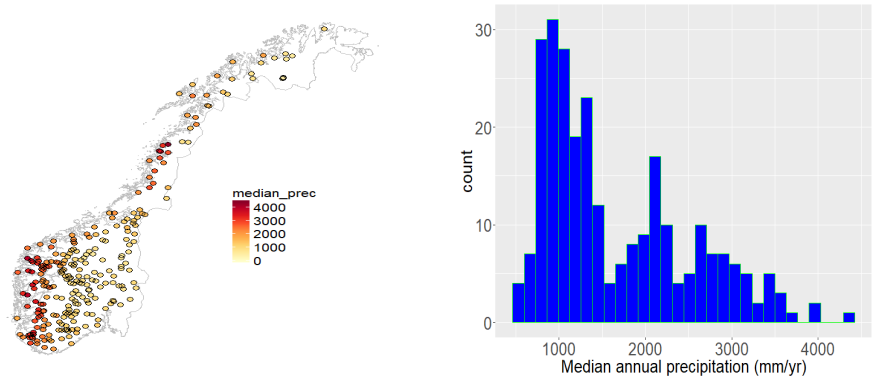


Figure 2.17: Scatter plot of median annual runoff plotted against the spatial dependency parameter (avg_ 5).



(a) Map of observed precipitation.

(b) Histogram of observed precipitation.

Figure 2.18: Map and histogram of observed precipitation (mm/yr).

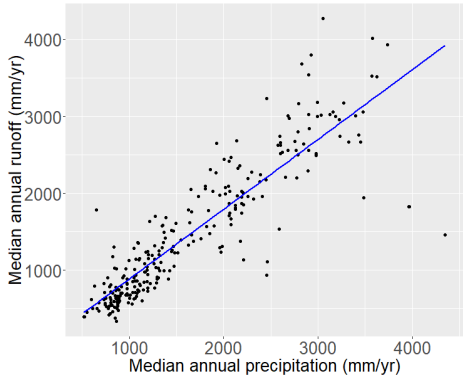


Figure 2.19: Scatter plots of observed median annual runoff vs. observed median annual precipitation.

Chapter 3

Background

In this chapter we present important background theory allowing us to construct statistical models for predicting runoff. We first present an introduction to linear mixed models, followed by the LGMs which are hierarchical Bayesian linear mixed models allow us to include a random field. Furthermore we explain random forest models and how these are built. We also present the most important parts of our spatial model, this comprises Gaussian random fields (GRFs), Gaussian Markov random fields (GMRFs) and the stochastic partial differential equation (SPDE) which links GRFs with GMRFs.

In the appendix we give an overview of Integrated Nested Laplace Approximation (INLA) approach for approximating the posterior marginals of the GRF, and allows fast inference and predictions for a LGM.

3.1 Linear Mixed Models

With linear models, multiple linear models and linear mixed models we assume that the response (runoff) can be modeled as a linear response of one or more explanatory variable (catchment characteristics, avg_5 and precipitation). We now present the basic theory for linear models with their most important assumptions. We introduce the linear models based on Fahrmeir et al. (2013) and refer to this work for further details.

Multiple linear models are an extension of the simple linear model presented in **eq. 2.1**. For our multiple linear model we assume that the observations of runoff $y_1, ..., y_i$ are independent and that that the distribution of runoff y_i depends on j continuous explanatory variables $x_{1,i}, ..., x_{j,i}$ where $j = 1, ..., m$. For $i = 1, ..., n$ observations of runoff we can can express the ith observation as

$$y_i = \beta_0 + \beta_1 x_{i,1} + ... + \beta_m x_{i,m} + \epsilon_i \tag{3.1}$$

where the random term $\epsilon_1, ..., \epsilon_n$ are assumed to be i.i.d. with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$. In **eq. 3.1** we also have that β_0 is the intercept, and $\beta_1, ..., \beta_m$ are the unknown coefficients for the explanatory variables $x_{i1}, ..., x_{im}$. **Eq. 3.1** can be written in matrix

notation as

$$\mathbf{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{3.2}$$

where $\mathbf{y} = [y_1, ..., y_n]$ are the vector of our dependent variables, $\mathbf{X}$ is a design matrix where the first row are ones corresponding with the intercept and the other rows are corresponding to the independent explanatory variables, the vector $\boldsymbol{\beta} = [\beta_0, ..., \beta_m]$ are the intercept and the unknown coefficients, and the vector $\boldsymbol{\epsilon} = [\epsilon_1, ...\epsilon_n]$ are our i.i.d. random errors with mean $\mathbf{0}$ and covariance $\sigma^2\mathbf{I}$.

When we extend our multiple linear models into a linear mixed model we incorporate some random effect in addition to the coefficient $\beta_1, ..., \beta_m$, into our model. A linear mixed model for the ith runoff observation can be expressed as

$$y_i = \beta_0 + \beta_1 x_{i,1} + ... + \beta_m x_{i,m} + \gamma_i + \epsilon_i \tag{3.3}$$

where γ_i is some random effect. If the response y_i is Gaussian distributed we make the assumption that the random effect γ_i is i.i.d. with mean 0 and variance τ^2. We can then interpret ϵ_i and γ_i as an unobserved process.

In matrix notation **eq. 3.3** becomes

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{3.4}$$

where we have that the random effect vector $\boldsymbol{\gamma} = [\gamma_1, ..., \gamma_n]$ is i.i.d. with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. The covariance matrix for two observations i and k is $\Sigma_{i,k} = \text{Cov}(\tau_i, \tau_k)$, $\text{Cov}(\tau_i, \tau_k)$ is a covariance function and specifies the correlation structure between two observations. The model in **eq. 3.3** induce a marginal correlation structure between the observations $\mathbf{y}$, such that $\mathbf{y}$ are conditionally independent given the unobserved process that ϵ_i and γ_i represents. Based on the assumption of i.i.d normal distributed unobserved process, we have the marginal distribution of $\mathbf{y}$ is

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I} + \tau^2\mathbf{J}). \tag{3.5}$$

The linear mixed model in **eq. 3.3** can be extended to Bayesian linear mixed models, where we assign a prior distribution to the random term $\boldsymbol{\epsilon}$ and the coefficients $\boldsymbol{\gamma}$.

3.2 Hierarchical models and latent Gaussian models

We now briefly introduce the Bayesian hierarchical models and the latent Gaussian models (LGMs). LGMs are hierarchically structured regression models where we use a Bayesian inference approach for computational benefits. This section is based on Gelfand et al. (2010) and Rue et al. (2009).

For Bayesian Inference we use a Bayesian hierarchical model of three layers, which consists of an observation model, a process (latent) model and a parameter model. The idea comes from basic probability theory where the joint distribution of some random variables can be broken up into conditional and marginal distributions, e.g. $P(A, B, C) = P(A|B, C)P(B|C)P(C)$ (Gelfand et al., 2010).

At the top level of hierarchy we have the observation model, for this we have that the dependent variables $\mathbf{y}$ are conditional dependent on a unobserved process $\boldsymbol{\eta}$ and some parameters $\boldsymbol{\theta_\tau}$. The observation model or observation likelihood can be expressed as $\pi(\mathbf{y}|\boldsymbol{\zeta}, \boldsymbol{\theta}_\tau) = \prod_{i=1}^{n} \pi(y_i|\eta_i, \boldsymbol{\theta_\tau})$.

At the second stage of our Bayesian hierarchy we have a model that describes the unobserved process $\boldsymbol{\eta}$. The unobserved process is a linear mixed model expressed as

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\zeta} \tag{3.6}$$

where $\mathbf{X}\boldsymbol{\beta}$ is the fixed effect and $\boldsymbol{\zeta}$ is a random field. The unobserved process is assigned a Gaussian distribution $\pi(\boldsymbol{\eta}|\boldsymbol{\theta}_\kappa)$ where $\boldsymbol{\theta}_\kappa$ is an vector with related parameters. In Gelfand et al. (2003) it is discussed how one can include spatially varying coefficients in the observation model, such that the explanatory variables included in the model account for local variations within the study area.

At the last stage we have the parameter model $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]$ that we assign some prior distribution.

3.3 Random forest

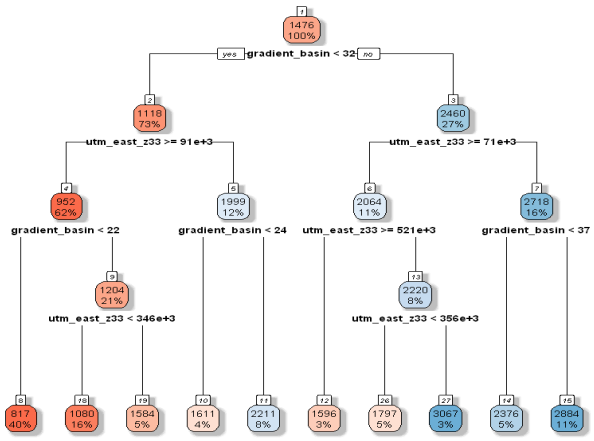
Random forest is a forest that consist of many decision trees. This section provides an overview of what a decision tree and a random forest is. For the random forest we refer to explanatory variables (catchment characteristics, avg_5 and observations of precipitation) as features or predictors and the dependent variable (runoff) is referred to as response. James et al. (2013) states that linear regression outperforms a regression tree if the relationship between features and the response is well approximated by a linear model. For a non-linear and complex relationship between the features and the response, a regression tree is likely to outperform a liner regression. This is much of our motivation for using random forest, as seen in section 2.5.2, most catchment characteristic does not have a linear relationship with median annual runoff.

In this section we introduce the regression tree, after this we explain the method of bagging which is used for bootstrapping our data and then we present the random forest which are a result of regression trees and bagging. We also introduce the partial dependency plots that allow us to visualise the marginal effect of one or more features on the predicted runoff.

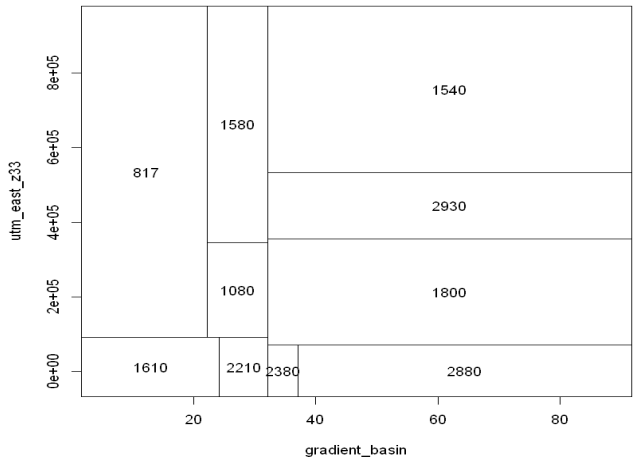
3.3.1 Regression trees

First out we start by explaining a regression tree, based on James et al. (2013). A regression tree is a tree-based method where we stratify the predictor space into two or more regions. To understand what a regression tree is we start out with an example based on the data in this thesis, where we would like to predict runoff based on some catchment characteristics. For simplicity we only use gradient basin and UTM east coordinates. After the example we explain how it is built.

An example of a regression tree is illustrated in **Fig. 3.1a** which is fitted to our data. In the regression tree a series of splitting rules is illustrated, from top of the tree to the



(a) An example of a regression tree for predicting runoff.



(b) The partition from the regression tree.

Figure 3.1: At left we have an example of a regression tree for a small sample of our data. The left-hand branches corresponds to $gradient\_basin < 32$ and the right-hand branches corresponds to $gradient\_basin \geq 32$. To the right we have an illustration of the partition of our data from the regression tree.

bottom. The top split assigns the runoff observed having gradient basin < 32 m/km to the left-hand branch. The predicted runoff for these catchments are given by the mean response value for the catchments within our data that has gradient basin < 32 m/km. For

those catchments we have a mean runoff of 1118 mm/yr. At the right-hand side we assign the catchments with gradient basin $\geq$ 32 m/km. This group is further subdivided by UTM east coordinates, and is continued until a stopping criteria is reached. In this case we stop growing the tree if the relative error does not improve more than 0.05. The result of our tree is plotted in **fig. 3.1b**, where we see that the tree stratifies the predictor space into ten regions. These regions are called terminal nodes or leaves of the tree.

We now explain how a regression tree is built. The tree is grown by first dividing our predictor space into two regions R_1 and R_2 where we have a mean response of 1118 mm/yr for R_1 and 2460 mm/yr for R_2, such that for a given response X = x, where $x \in R_1$ we predict runoff of 1118 mm/yr. The goal is to find regions $R_1, ..., R_j$ that minimize the mean squared error (MSE), which gives us

$$\frac{1}{J}\sum_{j=1}^{J}\sum_{i\in R_j}(y_i - \hat{y}_{R_j})^2 \tag{3.7}$$

here $\hat{y}_{R_j}$ is the mean response for the observations within the region R_j.

The partitioning is done by an approach called recursive binary splitting. Where the splitting begins at the top of the tree and the successively splits the predictor space. This is called a greedy approach because in each step of the process of building a tree each split is chosen as the best split, rather than looking into the future and look at what results in a better tree in some future step. By best split we mean the split that reduces the MSE the most, it is also possible to use other evaluation measures. Recursive binary splitting is performed as follows: First selecting a predictor X_j and a cut point s so that the splitting of the predictor space falls into two regions $X|X_j < s$ and $X|X_j \geq s$ leading to the larges possible reduction of MSE. Then for all the predictors $X_1, X_2, ..., X_j$, and all the possible values of the cut point, s, the predictor and cut points are considered and whichever combination that results in the lowest MSE is chosen. In mathematical form this is defined as,

$$R_1(j,s) = X|X_j < s \quad \text{and} \quad R_2(j,s) = X|X_j \geq s, \tag{3.8}$$

where we wish to find the j and s minimising

$$\frac{1}{J}\sum_{i:x_i\in R_1(j,s)}(y_i - \hat{y}_{R_1})^2 + \frac{1}{J}\sum_{i:x_i\in R_2(j,s)}(y_i - \hat{y}_{R_2})^2 \tag{3.9}$$

here $\hat{y}_{R_1}$ and $\hat{y}_{R_2}$ are the mean response in $R_1(j,s)$ and $R_2(j,s)$ respectively. This is done for each leaf on the tree until a stopping criteria is reached. When the regions have been defined we can predict the response (runoff) for a given observation.

Trees are easy to explain and visualise, but they tend to be sensitive to small changes in the data and therefor also poor as predictors. The average observations from several trees reduces the variance of our prediction.

3.3.2 Bagging

Bagging, bootstrap aggregation, is a method used for reducing variance of our predictions from regression tree. First introduced by, Breiman (1996). It is the process of randomly

picking observations from training data and making several subsets of the training data, with replacements. Some observations from the original training data may be observed several times. By making a separate tree for each of these subsets of the training data, we get equally many predictions as there are subsets of the training data. Averaging the prediction from each tree creates an overall predicted value. And in the process of bagging the training data, about 2/3 of the data is used. The 1/3 that is not used in each tree, are used to calculate an out-of-the-bag (OOB) prediction that can be obtained from the average of all the trees that does not use that observation. This way we can create an OOB error, and for a sufficiently large number of trees, the OOB error is virtually equivalent to the leave-one-out-cross-validation-error (James et al., 2013).

By applying bagging to a regression tree it becomes much harder to visualise the procedure, but in return it increases the accuracy. To get an overview of what is happening in the case of bagging regression trees, the decrease or increase in MSE can be recorded as new features are introduced in a tree, and then averaged over all the trees. Large increase in MSE indicates more important features.

In bagging all original features are considered at every split in all the trees, such that a broad part of the trees use its strongest predictor at the top. This creates correlation between the bagged trees, and the reduction of variance is not optimal. Therefor methods that force trees to not be correlated has been developed.

3.3.3 Random forest

Random forest is like bagging, but for random forest we do not only subset the data, we also subset the features to choose at each split. Random forest was introduced by Breiman (2001), and we refer to his work in the following section.

Random forest grows a number of decision trees where we at each leaf (split) within each tree only consider a random subset n of our predictors (features) and at the next split a new random sample of n predictors are considered. For each tree we are only considering a random sample of our original data set, e.g. using bagging. This method makes the trees less correlated and thus make our predictions more accurate.

3.3.4 Partial dependency plots

With partial dependency plots we are able to visualise the marginal effect of one or more features on the predicted runoff. It was first introduced by Friedman (2001), and the following is based on his work.

With partial partial dependency plots (PDPs) we visualise the partial dependence of the response $\hat{y} = \hat{F}(\mathbf{x})$ on a small selected subset of our features (explanatory variables). If $\mathbf{x}_l$ is the chosen subset of size l from the features $\mathbf{x}$,

$$\mathbf{x}_l = [x_1, ..., x_l] \subset [x_1, ..., x_n] \tag{3.10}$$

and $\mathbf{x}_{-l}$ is the complimented subset such that

$$\mathbf{x}_{-l} \cup \mathbf{x}_l = \mathbf{x} \tag{3.11}$$

we have that $\hat{F}(\mathbf{x})$ depending on the features from both subsets,

$$\hat{F}(\mathbf{x}) = \hat{F}(\mathbf{x}_l, \mathbf{x}_{-l}). \tag{3.12}$$

By conditioning on specific variables in $\mathbf{x}_{-l}$ we are able to consider $\hat{F}(\mathbf{x})$ as a function of the chosen subset $\mathbf{x_l}$,

$$\hat{F}(\mathbf{x}_l) = \hat{F}(\mathbf{x}_l|\mathbf{x}_{-l}). \tag{3.13}$$

Then the partial dependence of $\hat{F}(\mathbf{x})$ can be expressed as

$$\bar{F}_l(\mathbf{x_l}) = E_{z_{-l}}[\hat{F}(\mathbf{x})] = \int \hat{F}(\mathbf{x}_l, \mathbf{x}_{-l}) p_{-l}(\mathbf{x}_{-l} d\mathbf{z}_{-l} \tag{3.14}$$

where $\bar{F}_l(\mathbf{x_l})$ is the average effect of the chosen feature subset $\mathbf{x}_l$, and $p_{-l}(\mathbf{x}_{-l}$ is the marginal probability density of $\mathbf{x}_{-l}$.

3.4 Gaussian spatial models

Based on our main learning's of chapter 5.4 we now introduce an approach for modelling runoff with a spatial model. Our spatial models are reflected by an assumption stated by W. R. Tobler, in 1970, which was *everything is related to everything* (Tobler, 1970). This is known as the first law of geography, and underline our motivation for using a random field for modelling runoff.

For construction of spatial models we present the important background theory. First we introduce GRFs followed by the SPDE which allows us to link GRFs with GMRFs. Next we introduce GMRFs, latent Gaussian models and finally INLA which allows fast inference and prediction of our spatial models.

3.4.1 Gaussian processes and Gaussian random fields

Now an introduction of GRFs is presented, based on Cressie (1993) and Lindgren and Rue (2011). The GRFs allow use to account for spatial correlation and will be used as the random field in the second stage of our LGM (see section 3.2 and **eq. 3.6**).

We assume that $[\gamma(\mathbf{s}) : \mathbf{s} \in D]$ is a realization of a random field where D is a fixed subset of $\mathbb{R}^d$, our spatial dimension is $d = 2$. The random field is a GRF if all finite collections of $\boldsymbol{\gamma}$ are jointly Gaussian distributed, e.g.

$$[\gamma_1, ..., \gamma_n] \sim \mathcal{N}_n(\mu, \boldsymbol{\Sigma}) \quad n \geq 1 \tag{3.15}$$

where $\mathcal{N}_n$ the multivariate normal distribution with mean μ and covariance $\boldsymbol{\Sigma}$. The random field $\gamma(\mathbf{s})$ for the $i = 1, ...n$ locations $s_1, ..., s_n$ will be denoted as $\boldsymbol{\gamma} = [\gamma_1, ..., \gamma_n]$. If the covariance $\boldsymbol{\Sigma}$ is only a function of the relative position of two locations it is stationary, meaning that the distribution of $(\gamma(\mathbf{s_1}), ..., \gamma(\mathbf{s_n}))$ is the same as for $(\gamma(\mathbf{s_1}+z_1), ..., \gamma(\mathbf{s_n}+z_j))$. If the covariance only depends on the Euclidean distance between two locations it is

isotropic. A stationary and isotropic GRF then has

$$E[\gamma_i] = \mu$$
$$Var[\gamma_i] = \tau^2$$
$$Corr[\gamma_i, \gamma_k] = \rho(|\mathbf{s}_k - \mathbf{s}_i|)$$

for any location s_i and s_k, where $|\cdot|$ is the Euclidean distance.

Due to the partial differential equation approach that is presented later in this chapter, we have chosen the Matérn covariance function to construct the covariance matrix Σ, which specifies the dependency structure of the GRF. The stationary and isotropic Matérn covariance function is given by

$$\text{Cov}(s_i, s_k) = \frac{\sigma^2}{\Gamma(v)2^{v-1}}(\kappa|s_i - s_k|)^v K_v(\kappa|s_i - s_k|) \quad v > 0,\ \kappa > 0 \tag{3.16}$$

where σ^2 is marginal variance, $\Gamma(\cdot)$ is the gamma function, κ is a scale parameter, v is a shape parameter and K_v is the modified Bessel function of second kind and order v.

The range of the field is defined as

$$\rho = \frac{\sqrt{8v}}{\kappa} \tag{3.17}$$

and tells us at which distance between two observations it becomes approximately independent. In our work we let the scale parameter $v = 0$ and thus define the range as

$$\rho = \frac{\sqrt{8}}{\kappa} \tag{3.18}$$

and we use this for further work.

For inference GRFs have computational costs of $O(n^3)$, where n is the dimension of our covariance matrix.

3.4.2 The stochastic partial differential equation approach to spatial modeling

The following section is based on Lindgren and Rue (2011) and Lindgren and Rue (2015). In Lindgren and Rue (2011) it was shown that a GRF with Matérn covariance function can be expressed as a solution of a SPDE, where the SPDE is expressed as following

$$(\kappa^2 - \Delta)^{\alpha/2}(\tau\boldsymbol{\gamma}) = W \tag{3.19}$$

here we have that, $\kappa > 0$ is a scaling parameter, Δ is the two dimensional Laplacian defined as $\Delta = \sum_i^d \frac{\delta^2}{\delta x_i^2}$, α is a smoothness parameter, τ is a parameters controlling the variance, $\boldsymbol{\gamma}$ is a GRF and W is spatial Gaussian white noise with unit variance. κ is the same as for the Matérn covariance function in **eq. 3.18**, the parameter v from **eq. 3.18** is linked to the SPDE through

$$v = \alpha - \frac{d}{2} \tag{3.20}$$

and also the marginal variance σ^2 from **eq. 3.18** is related to the SPDE through

$$\sigma^2 = \frac{\Gamma(v)}{\Gamma(\alpha)(4\pi)^{d/2}\kappa^{2v}\tau^2} \tag{3.21}$$

For the range in **eq. 3.18** we specified that $v = 0$ for our work, further we also let $\alpha = 1$ and $d = 2$ and we can thus define the marginal variance as

$$\sigma^2 = \frac{1}{\sqrt{4\pi}\tau\kappa} \tag{3.22}$$

which we use for further work.

Using a finite element method we can approximate the solution of the SPDE in **eq. 3.19**, such that we can express our GRF as $\boldsymbol{\gamma}$ through a basis function representation. Where the domain of the GRF is discretizied into a triangular mesh (see **fig. 4.1** for a illustration of our mesh) with l mesh nodes and l basis functions ψ_o, e.g.

$$\boldsymbol{\gamma} = \sum_{o=1}^{l} \psi_o w_o \tag{3.23}$$

where $\mathbf{w}_o$ are Gaussian distributed weights with mean $\mu = 0$ and precision matrix $Q^{-1}(\tau, \kappa)$. The basin function ψ_o is defined such that $\psi_o = 1$ at vertex o and 0 at all other vertices. **Eq. 3.23** can be interpreted in the sense that the weights determine the value of the filed in the vertices and the interior of the triangles are determined by linear interpolation. The joint distribution of the weights determines the full distribution for the continuously indexed solution.

The precision matrix $Q^{-1}(\tau, \kappa)$ for the weights $\mathbf{w_o}$ is defined such that $\boldsymbol{\gamma}$ is continuously indexed. For the smoothness parameter $\alpha = 2$, we have that Q is defined as

$$Q(\tau, \kappa) = \tau^2(\kappa^4\mathbf{C} + 2\kappa^2\mathbf{G} + \mathbf{G}\mathbf{C}^{-1}\mathbf{G}) \tag{3.24}$$

here $C_{ij} = (\psi_i\psi_j)$, $G_{ij} = (\nabla\psi_i\nabla\psi_j)$ for $i = 1, ..., n$ and $j = 1, ..., m$. We use a common parameterization of the precision matrix, where $\log(\tau) = \theta_\tau$ and $\log(\kappa) = \theta_\kappa$. Which gives the following presentation of Q

$$Q(\theta_\tau, \theta_\kappa) = \exp(2\theta_\tau)[\exp(4\theta_\kappa)\mathbf{C} + 2\exp(2\theta_\kappa)\mathbf{G} + \mathbf{G}\mathbf{C}^{-1}\mathbf{G}] \tag{3.25}$$

now we have that $Q(\theta_\tau, \theta_\kappa)$ as a sparse matrix.

3.4.3 Gaussian Markov random fields

Now we introduce Gaussian Markov random fields (GMRFs). We will represent our GRFs as GMRFs as allow us make faster inference and predictions for our spatial models. This section is based on the work of Rue et al. (2009) and Rue and Held (2005) and we refer to this work for further details.

With the SPDE approach from **eq. 3.23** we can represent our GRF from section 3.4.1 as a GMRF $\boldsymbol{\gamma}$ with mean μ and precision matrix $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$ characterized by a Markov

property. For the Markov property we have that the GMRF $\boldsymbol{\gamma} = [\gamma_1, ..., \gamma_n]$ for the locations $i = 1, ...n$, where the vector $\boldsymbol{\gamma}_-$ is the vector $\boldsymbol{\gamma}$ not containing element i. Where the conditional distribution of γ_i for each location i only depends on a set of neighbors δ_i. This Markov property is reflected in the sparse precision matrix $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$. $\mathbf{Q}$ is called a sparse precision matrix as it contains many zeros due to the conditional distribution of γ_i.

The sparse precision matrix $\mathbf{Q}$ allows for a large reduction in computational cost of compared to the computational cost of the GRF (see section 3.4.1), as the computational cost of a GMRF is typically of $O(n^{3/2})$.

3.4.4 Integrated nested Laplace approximation

For Bayesian inference of LGMs as presented in section 3.2 we use the Integrated nested Laplace (INLA) approach. Traditionally Markov chain Monte Carlo (MCMC) sampling was used to do inference for models such as the LGMs (see Robert and Casella (2004)). INLA was presented by Rue et al. (2009), where they argue that for a given the computational cost, INLA approach outperforms MCMC. Thus our motivations for using INLA is due to the high computational speed compared to the MCMC.

As we are within the Bayesian world the goal is to approximate the marginal posterior distribution of the unobserved process $\pi(\eta_i|\theta_\kappa)$ for $i = 1, ..., n$ and the posterior distribution of the parameter model $\pi(\theta)$. We now present the LGMs suited for INLA based on Lindgren and Rue (2015) and Rue et al. (2009).

For our LGM we must have some requirements fulfilled. Further we need our unobserved process η to be a GMRF, such that numerical methods for sparse matrices can be used. Secondly our parameter model $\pi(\theta)$ should be small, meaning number of hyperparameters θ should not be much larger than approximately 5.

As stated, the goal of Bayesian inference are the marginal posterior distributions for each element of the latent model and each element of the parameter model. They are computed with the following integrals

$$\pi(\eta_i|\mathbf{y}) = \int p(\eta_i, \theta|\mathbf{y})d\theta = \int p(\eta_i|\theta, \mathbf{y})p(\theta|\mathbf{y})d\theta \tag{3.26}$$

$$p(\theta_j|\mathbf{y}) = \int p(\theta|\mathbf{y}d\theta_{-\mathbf{j}} \tag{3.27}$$

these needs integrals to be computed. For our spatial models we have that likelihood $\mathbf{y}$ is Gaussian distributed. The integrals in **eq. 3.26** and **eq. 3.26** are determined by numerical approximation that er described in Rue et al. (2009).

3.5 Evaluation measures

We now present the different evaluations used to compare the predictive power of our models. By evaluating the predictive power we explore our models ability to make accurate predictions $\hat{\mathbf{y}}$.

3.5.1 Coefficient of determination

We use the coefficient of determination R^2 to compare our linear models that we present in chapter 4.1. The coefficient of determination R^2 was first introduced in Wright (1921), and for presenting the coefficient of determination R^2 we refer to Fahrmeir et al. (2013).

The coefficient of determination R^2 is closely related to the correlation coefficient r that we introduced in 2.5.1, where we used the correlation coefficient for exploring the linear association between the dependent variable y_i and a independent explanatory variable x_i for $i = 1, ..., n$. In the case of a simple linear regression model as described in **eq. 2.1** the coefficient of determination R^2 corresponds to the squared correlation coefficient r, e.g. $R^2 = (r)^2$. For multiple linear regression models as described in **eq. 3.1** the coefficient of determination R^2 is the squared correlation coefficient between the observations $\mathbf{y} = [y_1, ..., y_n]$ and the minimum mean squared error (MSE) predicted values $\hat{\mathbf{y}} = [\hat{y}_1, ..., \hat{y}_n]$ which implies that $R^2 = (r)^2$.

R^2 represents the proportion of variance of runoff that is explained by the explanatory variables, and is defined as

$$R^2 \quad = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^{n}\hat{\epsilon}_i}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \tag{3.28}$$

where $\bar{\mathbf{y}}$ is the average of the minimum MSE predicted runoff $\hat{\mathbf{y}}$, and $\hat{\epsilon}$ is the minimum MSE estimated residuals. The value of R^2 ranges between 0 and 1, where a high value indicates a better fit.

The R^2 increase in value as more explanatory variables are included in the model while the adjuster R^2 (also known as corrected coefficient of determination) accounts for the number of explanatory variables included in the models. The adjusted R^2 are defined as follows

$$\bar{R}^2 = 1 - \frac{m-1}{m-p}(1 - R^2) \tag{3.29}$$

where m is the number of explanatory variables and p is the p-value. The adjusted R^2 is supposed to penalize for increased number of explanatory variables, but for tests scores larger than 1 it starts to increase and is therefor not a preferred evaluation measure.

3.5.2 Root mean square error

Measuring predictive performance of all models presented in this thesis are done with root mean square error (RMSE), which measures the squared difference between predicted $\hat{y}_i$ runoff and observed y_i runoff. The difference between predicted $\hat{y}_i$ runoff and observed y_i runoff are what we refer to as residuals. The RMSE indicates accurate predictions $\hat{y}$ by a low value in the unites of the dependent variable (runoff).

The RMSE is defined as

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{3.30}$$

3.5.3 Continuous ranked probability score

In addition to the RMSE we use the continuous ranked probability score (CRPS) for measuring the predictive performance of our models. With CRPS we are able to evaluate the whole posterior predictive distribution of our predicted value $\hat{y}_i$ (Ingebrigtsen et al., 2015). When we account for the whole posterior predictive distribution we asses both the sharpness and the precision of our predicted value $\hat{y}_i$. With sharpness we refer to small standard deviation of posterior distribution, while precision refers to the accuracy of the predicted outcome $\hat{y}_i$ when we compare it to the observed value y_i. An accurate prediction is indicated by a low value of mean CRPS in the units of the dependent variable y_i.

Gneiting and Raftery (2007) defines the CRPS as follows,

$$CRPS(F, y) = \int_{-\infty}^{\infty} (F(u) - 1(y \geq u)^2 du \tag{3.31}$$

where F is the predicted cumulative distribution function of the predicted value $\hat{y}_i$ and y_i is the observed value.

3.6 Evaluation schemes

3.6.1 Leave-one-out cross validation

The models we use for this thesis are tested through a evaluation scheme named leave-one-out cross-validation (LOOCV). LOOCV leaves out one observation y_i from the original data with $i = 1, ..., n$ observations, and the remaining $(n-1)$ observations $\mathbf{y}_-$ are used to construct a model. With the constructed model we obtain a prediction $\hat{y}_i$ of the observation y_i left out. Each predicted value $\hat{y}_i$ is then evaluated in terms of some evaluation metric. This procedure is repeated n times, such that we have n predicted values $\hat{y}$ and n scores from some evaluation measure.

To compare the predictive performance of the models built in this thesis we use the two evaluation metrics RMSE and CRPS presented in section 3.5. We use both the n individual RMSE and CRPS scores, and the mean RMSE ($\overline{RMSE}$) score and the mean CRPS ($\overline{CRPS}$) score for comparison.

The $\overline{RMSE}$ and $\overline{CRPS}$ are calculated as follows

$$\overline{RMSE} = \frac{1}{n} \sum_{i=1}^{n} RMSE_i \tag{3.32}$$

$$\overline{CRPS} = \frac{1}{n} \sum_{i=1}^{n} CRPS(F_i, y_i). \tag{3.33}$$

Chapter 4

Models for prediction of median annual runoff

We now present the models used to predict runoff in this thesis, and also how we have inference for such models. First we do an overview of the multiple linear models, and then an overview of our random forest models. Next we present how we construct our spatial models (LGM) with the SPDE approach.

Our linear regression models and random forest models have been created as a tools for exploring the relationships between our catchment characteristics and observed runoff, thus all available catchment characteristics are included. We also investigate how spatial dependency in terms of how average neighbor runoff (avg_5) and observations of precipitation influence our predictions.

4.1 Multiple linear regression method

We have created three different linear models, the first model is named lm_c and does only include the catchment characteristics $x_{j,i}$ (listed in **tab. 2.1**). Next we have created a model named lm_cn which includes our catchment characteristics $x_{j,i}$ plus average neighbour runoff avg_5$_i$. The last model created is the model lm_cnp, which contains all the catchment characteristics $x_{j,i}$, average neighbour runoff avg_5$_i$ and precipitation p_i.

The simple multiple linear regression model we use are as described in section 3.1, where we have

$$y_i = \beta_0 + \beta_1 x_1 + ... + \beta_m x_{m,i} + \epsilon_i \tag{4.1}$$

our response variable y_i is median annual runoff at catchment s_i where $i = 1, ..., 266$, and that β_0 is the intercept and $\beta_1, ..., \beta_m$ are the unknown coefficients of our explanatory variables $x_1, ..., x_m$. The different explanatory variables x_m for $j = 1, ..., m$ depend on which model we are considering. For the random term ϵ_i we make the same assumptions as we specified in section 3.1.

4.1.1 Inference for linear regression

To estimate the explanatory variables β_m and the random term ϵ_i we use the minimum MSE. With the minimum MSE estimators we get the predicted median annual runoff at catchment s_i as

$$\hat{y}(s_i) = \hat{\beta}_0 + \hat{\beta}_1 x_1(s_i) + ... + \hat{\beta}_n x_n(s_i) + \hat{\epsilon}(s_i) \tag{4.2}$$

with this we are able to evaluate the predictive performance with RMSE and CRPS.

4.1.2 LM models

The different liner regression models used to predict median annual runoff are listed in the **tab. 4.1**.

	Model	Catchment characteristics	Average neighbor runoff	Precipitation
1	LM_c	YES	NO	NO
2	LM_cn	YES	YES	NO
3	LM_cnp	YES	YES	YES

Table 4.1: Table with the three linear models (LM) built for this thesis and what parameters they contain. Catchment characteristics $x_{i,j}$ are listed in **tab. 2.1**, average neighbor runoff avg_5$_i$ is the average runoff of the five closest neighbors (see section 2.6) and precipitation p_i is the observed median annual precipitation at each catchment (see section 2.7).

4.2 Random forest method

We have built three different random forest models, in the same manner as the linear models described in section 4.1. For our random forest models we did a manual search, for constructing the best possible random forest model. With a manual search we tested many combinations of how many features that each split should be allowed to select from, maximum number of leafs a tree should contain and how many trees the forest should consist of. We determined the best structure of combination in terms of what forest resulted in the best MSE.

As there are many possible combinations of how to build a tree, we have not been able to test all. The models we present are the models that was found to have the smallest MSE score after a process of testing some combinations. **Tab. 4.2** is presented with what was found as the best structure for a each random forest model.

4.2.1 RF models

The different random forest models used for predicting runoff are listed in **tab. 4.3**.

	Model	Max depth	Number of features	Number of trees
1	RF_c	20	5	500
2	RF_cn	12	10	500
3	RM_cnp	16	14	500

Table 4.2: Set up for the random forest models (RF) created by a manual search of possible combinations of number of leafs (max depth), number of features to select at each split (number of features) and number of trees that the forest consist of (number of trees).

	Model	Catchment characteristics	Average neighbor runoff	Precipitation
1	RF_c	YES	NO	NO
2	RF_cn	YES	YES	NO
3	RF_cnp	YES	YES	YES

Table 4.3: Table of the three random forest models (RF) built for this thesis and what parameters they contain. Catchment characteristics are listed in **tab. 2.1**, average neighbor runoff avg_5$_i$ is the average runoff of the five closest neighbors (see section 2.6) and precipitation p_i is the observed median annual precipitation at each catchment (see section 2.7).

4.3 Latent Gaussian model

The spatial models built for this thesis are based on the assumption that our observations are correlated and the result of exploration of our linear models (section 4.1) and random forest models (section 4.2). The main learning's of the linear models and random forest models can be found in section 5.4. We now construct our spatial models as the LGMs presented in section 3.2.

For this thesis we present four different spatial models. The first model, SP_r only contains a GRF. The second model, SP_rb, contains a GRF and the explanatory variable g_i which is the catchment characteristics gradient basin. The third model, SP_rbp, contains a GRF and the two explanatory variables g_i and precipitation p_i. For the fort model, SP_rbpc we have included a spatially varying coefficient such that we have the same construction as for SP_rbp, but now also allowing the coefficient of precipitation to vary spatially, this is denoted as $\tilde{\beta}_{j,i}$.

As introduced in section 3.4 our LGMs are hierarchical models of three levels. We now present the different models within each level.

Observation model

For our observation model we have Norway as our spatial domain D, such that $D \subset R^2$. We denote true runoff at the point location (centroid) $s_i \in D$ as y_i where $i = 1, ...n$. The observed runoff y_i is observed with some random error ϵ_i for location s_i. Thus the observation model for true runoff is

$$y_i = \eta_i + \epsilon_i \tag{4.3}$$

where the random error ϵ_i is assumed to be i.i.d. Gaussian distributed with mean 0 and variance (precision) τ_p^{-1}, and also independent of the unobserved process η_i. The observa-

tion model for true runoff in **eq. 4.3** is for the three model SP_r, SP_rb and SP_rbp, where we do not have any spatially varying coefficients.

For the model SP_rbpc we let the coefficient of precipitation vary within our spatial domain (Norway). For SP_rbpc we express the observation model for true runoff y^*_i at location s_i as

$$y^*_i = \eta^*_i + \epsilon^*_i \tag{4.4}$$

where we assume that the random error ϵ^*_i is i.i.d Gaussian distributed with mean 0 and precision τ_c^{-1}. The random error ϵ^*_i is also assumed to bee independent of the unobserved process η^*_i.

Process model

Our process model is assumed to model the true level of runoff, and expressed as

$$\eta_i = \beta_0 + \gamma_i \tag{4.5}$$

where β_0 is the intercept and γ_i is a stationary and isotropic GRF. **Eq. 4.5** is the process model of SP_r. For SP_rb we also include the explanatory variable g_i and for SP_rbp we include the two explanatory variables g_i and precipitation p_i, this gives the following process model for SP_rb

$$\eta_i = \beta_0 + \beta_1 g_i + \gamma_i. \tag{4.6}$$

For SP_rbp we have the following process model

$$\eta_i = \beta_0 + \beta_1 g_i + \beta_2 p_i + \gamma_i. \tag{4.7}$$

For SP_rbpc we include the spatially varying coefficient $\tilde{\beta}_i$, with the following process model

$$\eta^*_i = \beta_0 + \beta_1 g_i + \tilde{\beta}_i p_i + \gamma_i, \tag{4.8}$$

where we interpreter $\tilde{\beta}_i = \beta_2 + \beta_i$ as a spatially varying coefficient, where β_i is a spatially random adjustment at location s_i for the explanatory variable p_i, thus β_i is a GRF as γ_i. We can thus express the process model of SP_rbpc as

$$\eta^*_i = \beta_0 + \beta_1 g_i + \beta_2 p_i + \beta_i p_i + \gamma_i. \tag{4.9}$$

For further detail on spatially varying explanatory variables we refer to Gelfand et al. (2003).

SPDE approach

For expressing our GRF γ_i as a GMRF we use the SPDE approach described in section 3.4.2. When we use the SPDE approach we discreetly index the GRF γ_i into a mesh covering Norway as seen in **fig. 4.1**, this is the mesh that have been used for all our spatial models. Whit triangulation mesh in **fig. 4.1** our GRF γ_i can be expressed as

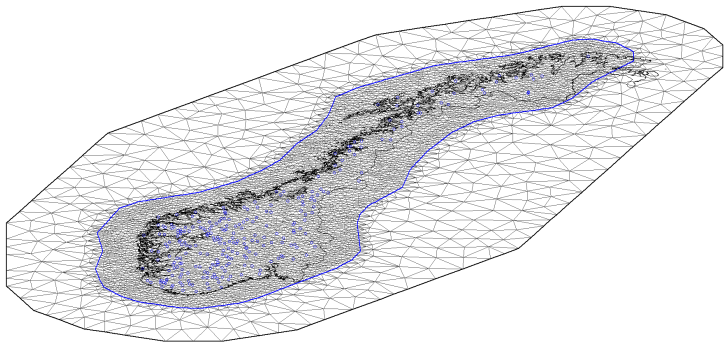


Figure 4.1: The mesh used in INLA for solving our SPDEs. Within the mesh we have added the border of Norway, and blue points locating our runoff observation locations. Our mesh has 4949 mesh nodes.

$$\gamma_i = \sum_{k=1}^{l} \psi_{k,i} w_k, \tag{4.10}$$

where $\psi_{k,i}$ is the basis functions constructed in the mesh with $k = 1, ..., l$ vertices and $l = 4949$. w_k is the approximation of γ_i at the k mesh nodes. In order to let the w_k be an GMRF it assigned a Gaussian distribution with 0 mean and the precision matrix $\mathbf{Q}^{-1}(\theta_{\tau,u}, \theta_{\kappa,u})$, which is expressed as

$$\mathbf{w} = [w_1, ..., w_k, ..., w_m]^T \sim N(\mathbf{0}, \mathbf{Q}^{-1}(\theta_\tau, \theta_\kappa)) \tag{4.11}$$

here $\theta_\tau = log(\tau)$ and $\theta_\kappa = log(\kappa)$, where θ_τ is linked to the range ρ from **eq. 3.18** and θ_κ is linked to the marginal variance σ^2 from **eq. 3.22**.

Inserting **eq. 4.10** into the process model for SP_r **eq. 4.5** we obtain the following,

$$\eta_i = \beta_0 + \sum_{k=1}^{l} \psi_{k,i} w_k \tag{4.12}$$

and the process model for SP_rb in **eq. 4.6** becomes

$$\eta_i = \beta_0 + \beta_1 g_i + \sum_{k=1}^{l} \psi_{k,i} w_k. \tag{4.13}$$

The process model for SP_rbp in **eq. 4.7** becomes

$$\eta_i = \beta_0 + \beta_1 g_i + \beta_2 p_i + \sum_{k=1}^{l} \psi_{k,i} w_k. \tag{4.14}$$

For expressing the process model η^*_i at locations s_i for SP_rbpc in **eq. 4.9** we first express our spatially random adjustment for precipitation (GRF) β_i as

$$\beta_i = \sum_{k=1}^{l} \delta_{k,i} u_k \tag{4.15}$$

where $\delta_{k,i}$ is the basins function constructed in the mesh from **fig. 4.1** with $k = 1, ..., l$ vertices and $l = 4949$. u_k is the approximation of β_i at the k mesh nodes, and by assigning it a Gaussian distribution with mean 0 and precision matrix $\mathbf{Q}^{-1}(\theta_{\tau,w}, \theta_{\kappa,w})$. Thus our spatially random adjustment for precipitation β_i is a GMRF. The process model of SP_rbpc can now be expressed as,

$$\begin{aligned} \eta^*_i &= \beta_0 + \beta_1 g_i + \beta_2 p_i \\ &\quad + p_i \sum_{k=1}^{l} \delta_{k,i} u_k + \sum_{k=1}^{l} \psi_{k,i} w_k \end{aligned} \tag{4.16}$$

where we have the GRF γ_i expressed as the GMRF described in **eq. 4.10**. And the spatially random adjustment β_i is expressed as a GMRF through **eq. 4.15**.

4.3.1 Inference for spatial models

Our spatial models are Bayesian linear mixed models which contains a GRF. For fast computation and inference we have used the INLA approach. For readability of this chapter we present the construction of our spatial models within the INLA framework in the appendix D.

From the spatial models we obtain the posterior distribution (probability distribution) of the different parameters within the model. From the posterior distribution we obtain the posterior mean through the minimum MSE estimator, which for the predicted runoff $\hat{y}_i$ from SP_r can be seen as

$$\hat{y}(s_i) = \hat{\beta}_0 + \hat{x}(s_i), \tag{4.17}$$

and it would be similar for all our spatial models which we have listed in **tab. 4.4**.

4.3.2 SP models

The different spatial models used to predict median annual runoff is listed in **tab. 4.4** below.

	Model	random spatial field	Gradient basin	Precipitation	SVC
1	SP_r	YES	YES	NO	NO
2	SP_rb	YES	YES	NO	NO
3	SP_rbp	YES	YES	YES	NO
4	SP_rbps	YES	YES	YES	YES

Table 4.4: The four SP models and what explanatory variables they contain, the random spatial field $\boldsymbol{\gamma}$, gradient basin g_i is the catchment characteristics listed in **tab. 2.1**, precipitation p_i is as described in section 2.1 and SVC is the spatially varying coefficients of precipitation $\tilde{\beta}_{j,i}$.

4.4 Software

In this section we will present the different packages used to build our models, do predictions of runoff, compare in terms of predictive performance and visually investigate.

For our linear models we have used a base R package named *stats* by R Core Team (2016). The *stats* package was also used for predictions with both our linear models and our random forest models. Plotting is done with the *ggplot2* package by, Wickham (2009).

To evaluate the models the RMSE and CRPS from LOOCV was calculated with the *Metrics* package and the *scoringRules* package respectively. The *Metrics* package is by Hamner (2012) and the *scoringRules* is by Jordan et al. (2016).

The random forest models where built with the *randomForest* package by Liaw and Wiener (2002). With the *randomForest* package we were also able to view variable importance plots. With the result of our *randomForest* model we were also able to view partial dependency plot with the R-package *pdp* by Greenwell (2016).

All of our spatial models have been built with the R-package *INLA* by Rue et al. (2009), *INLA* is available at *www.r-inla.org*. Within the INLA framework we use the SPDE approach which was presented by Lindgren and Rue (2011).

Chapter 5

Results initial model exploration

In this chapter we present the main results of the linear models and random forest models introduced in chapter 4 (see **tab. 4.1** and **tab. 4.1**). In the first section we present the results of the linear models and evaluate the performance and uncertainties of the explanatory variables included. In the next section we present the results of the random forest models, with focus on feature importance and the partial dependency of features (explanatory variables) within the models. Furthermore we compare and evaluated the predictive performance of our linear models and random forest models.

With the linear models and our random forest models we want to explore what catchment characteristics are most influential on the prediction of runoff. We also want to explore how the explanatory variables/features average neighbour runoff (avg_5) and precipitation, influence the estimated coefficients and predictive power of our models.

The main learning's from the linear model and the random forest will be summarised in the end of this chapter, the main learning's are used for construction of our spatial models.

5.1 Results linear model

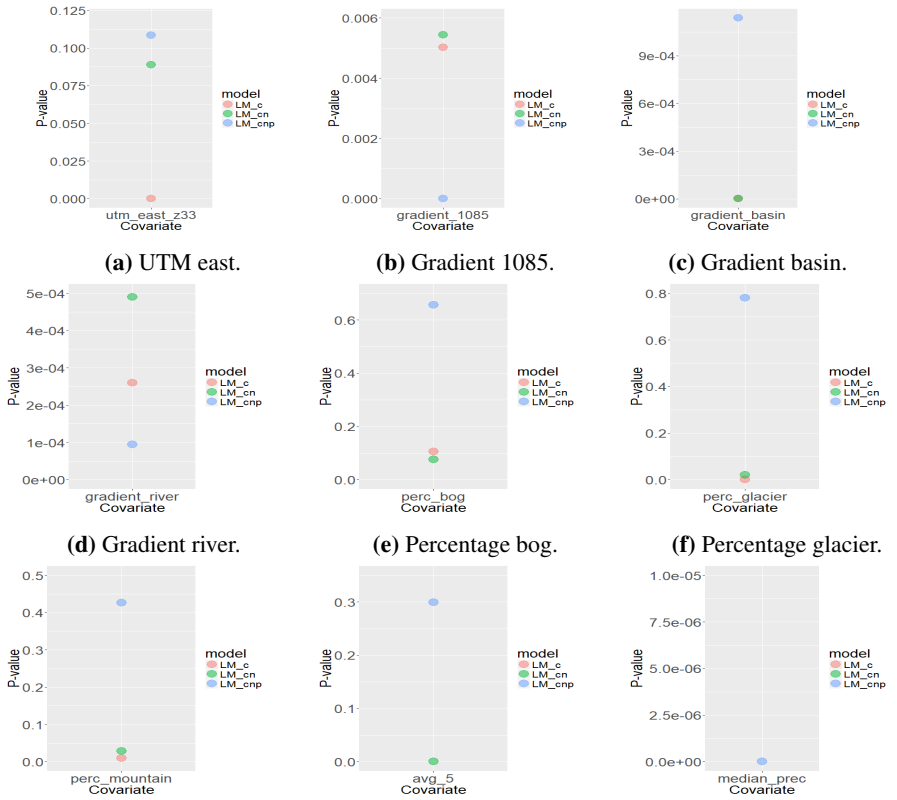
We first consider the linear model LM_c where the explanatory variables are the catchments characteristics presented in **tab. 2.1**. For LM_c the model estimates for the coefficients, together with their standard deviation is presented in **tab. E.1**. The model output of LM_c shows that the catchment characteristics that have a significant p-value on a 5% level are the coefficients of UTM east coordinates, gradient basin, gradient river and percentage of glacier.

Tab. E.3 are the same results for the model where the average of neighbours (avg_5) is included as a explanatory variable, LM_cn. If we compare the model estimates for the coefficients in **tab. E.1** of LM_c with the estimated for the coefficients of LM_cn in **tab. E.3** we find that the p-values change, and the coefficient for the UTM east coordinates are no longer significant and neither is the percentage of glacier. The coefficients for gradient basin and gradient river both remain significant in LM_cn, so is the coefficient for avg_5.

Tab. E.5 are the estimated coefficients for the models that also include precipitation LM_cnp. Here we find that the estimated coefficient for precipitation and the catchment characteristics gradient 1085 have a significant p-value, we also find that the p-value of the estimated coefficient of avg_5 is not significant.

Fig. 5.1 allow us to visualise the change in p-value for the estimated coefficients of our linear models. When we compare the p-values for the estimated coefficient of avg_5 in **fig. 5.1h**, we observe a difference between the two models LM_cn and LM_cnp. This change in p-value indicates that precipitation seem to account for much of the same spatial dependency that avg_5 accounts for. Further we also notice that the estimated coefficients have much larger p-values for LM_cnp than LM_cn.

Fig. 5.1 also illustrates how the p-values of the catchment characteristics decrease as we introduce information about either neighbouring catchments and precipitation. This indicates that the catchment characteristics are less influential in the linear model when we include some explanatory variables that describes the spatial dependency between our observation of runoff.



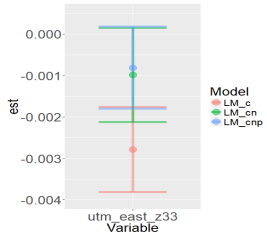
(a) UTM east. **(b)** Gradient 1085. **(c)** Gradient basin.

(d) Gradient river. **(e)** Percentage bog. **(f)** Percentage glacier.

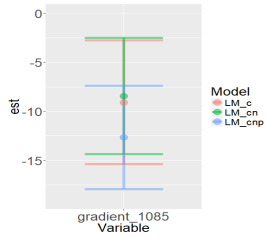
(g) Percentage mountain. **(h)** Spatial dependency. **(i)** Precipitation.

Figure 5.1: Plots of the p-values within the three linear models (LM_c, LM_cn and LM_cnp). The plots illustrates how the p-values of the coefficient for the explanatory variables change within the different linear models.

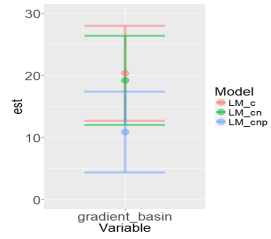
The estimated coefficients decrease in estimated value as avg_5 and precipitation are included, but there are not much change in their uncertainty. This is illustrated in **fig. 5.2**, where the significant (on a 5% level) estimated coefficients are plotted with their corresponding 95% confidence interval. The only estimated coefficient with a visibly smaller confidence interval when avg_5 and precipitation are included is the coefficient gradient river seen in **fig. 5.2d**.



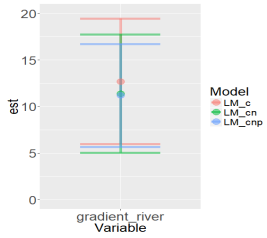
(a) UTM east.



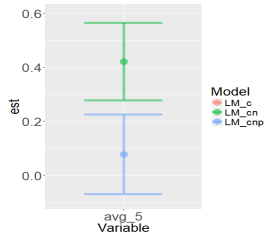
(b) Gradient 1085.



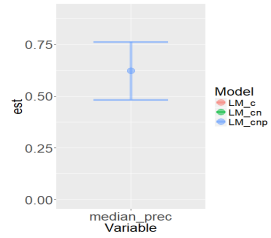
(c) Gradient basin.



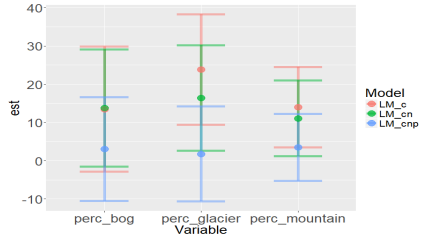
(d) Gradient river.



(e) Spatial dependency.



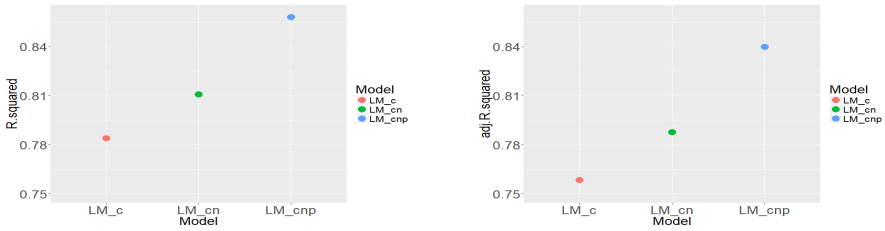
(f) Precipitation.



(g) Percentage bog, glacier and mountain.

Figure 5.2: 95% confidence interval for the estimated coefficients for the explanatory variables in the linear models LM_c, LM_cn and LM_cnp.

To illustrate the effect of the two explanatory variables avg_5 and precipitation for modeling runoff we can have a look at the plots with the coefficient of determination R^2 and the adjusted coefficient of determination R^2_{adj} in **fig. 5.3**. Both coefficient of determination R^2 and adjusted coefficient of determination R^2_{adj} illustrates an increasing value when the catchment characteristics no longer are the only covariates. Therefore it seems like our predictions of runoff will improve if we are able to include information about spatial dependency. We also notice that the adjusted coefficient of determination R^2_{adj} in **fig. 5.3b** is larger for LM_cnp than for LM_cn, indicating that precipitation increase the predictive performance of our linear models.



(a) Multiple R^2. **(b)** Adjusted R^2.

Figure 5.3: The multiple R^2 and the adjusted R^2 for the linear models.

5.2 Results random forest

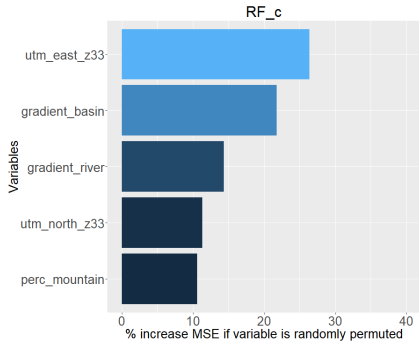
Random forest serves the same purpose as our linear models, further it enable us to explore the functional form (non-linear) and interactions between features.

The importance of the different features within our random forest models are illustrated in **fig. 5.4**. Here we can see how the importance of different features changes with the different models. For RF_c (with catchment characteristics as features) the most important feature are the UTM east coordinates followed by gradient basin. When the feature average neighbour runoff (avg_5) are included in RF_cn we observe that it is the most important feature, and also that the features which was most important for RF_c becomes less important. For RM_cnp precipitation accounts for more than 30% of the decrease in MSE, indicating that all other features have a much smaller impact on the model compared to LM_c and LM_cn.

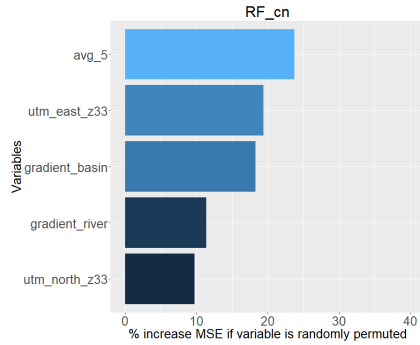
Further we explore the functional form of our random forest models by analysing the partial dependency plots of **fig. 5.5**. Here we have only included the features that seems to have the greatest importance based on the results of our variable importance plots seen in **fig. 5.4**. We observe that the marginal effect of the UTM east coordinates with a small/negative value have a large predicted runoff, and that this quickly decrease as we move further towards the east (larger value of UTM east, see map in **fig. 2.1**). The UTM north coordinates have a more complicated marginal effect on the predicted runoff. Where we in **fig. 5.5b** see that past $6.7*10^6$ meters increasing UTM north coordinates also gives an increase in predicted runoff. The gradient basin have an increasing predicted runoff with increasing gradient basin past 20 m/km, past 40 m/km the marginal effect is constant, and the marginal effect is similar for gradient river. For percentage of forest (**fig. 5.5e**) and percentage of mountain (**fig. 5.5f**) do both seem to have a linear trend with predicted runoff, for percentage of forest the marginal effect is decreasing and for percentage of mountain marginal effect is increasing. The feature describing neighbouring catchments have a linearly increasing relationship with the predicted runoff, and so does also the median precipitation.

What we notice from the marginal effect of our most important features in **fig. 5.5** is that they all seems to have an approximately linear marginal effect on predicted runoff.

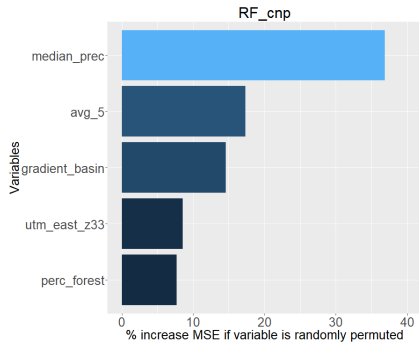
The partial dependency plots in **fig. 5.5** illustrate what we have observed from the variable importance plots in **fig. 5.4**. When the features avg_5 and precipitation are introduced in the random forest models, the marginal effects of the catchment characteristics



(a) Variable importance for RF_c.



(b) Variable importance for RF_cn.



(c) Variable importance for RF_cnp.

Figure 5.4: Variable importance for our random forest models. The top 5 most important features for RF_c, RF_cn and RF_cnp, importance is decided from what features result in the largest percent decrease in MSE.

are reduced. Comparing the marginal effect of the features contained in LM_c and LM_cn, in **fig. 5.5** observe how the marginal effect is larger within the RF_c model than in RF_cn. Further we observe that the marginal effect of our features in RF_cnp are even smaller than the marginal effect of the features in LM_cn.

The marginal effect of precipitation seem to dominate the random forest models based on the partial dependency plots in **fig. 5.5**. From the plot with variable importance (**fig. 5.4**) we also observe that precipitation also account for the largest decrease in MSE error. The marginal effect of avg_5 increase when precipitation is excluded from the model (RF_cn), and so does the variable importance. Which indicates that spatial dependency is sufficient for predicting runoff.

The random forest models that only contains catchment characteristics (RF_c) indicates that UTM east is the most important feature, and it is not a large difference between UTM east and gradient basin in terms of decrease in MSE. As UTM east coordinates defines how far east the observation is located, it seem like it is able to account for some of the spatial dependency that the features avg_5 and precipitation accounts for. This is further indicated by how UTM east perform in terms of importance for the two models

RF_cn and RF_cnp, where we in **fig. 5.5** observe how gradient basin are more important. Thus indicating that gradient basin is able to account for some other effect than the spatial dependency.

From the partial dependency plots in **fig. 5.6** we observe that the marginal effect of both gradient basin and avg_5 have local variations within Norway. While precipitation have an overall larger marginal effect, thus indicating that precipitation is not able to detect local variations as well as gradient basin and avg_5.

5.3 Predictive performance

	Model	mean RMSE	mean CRPS
1	LM_c	347.14	254.70
2	LM_cn	323.89	238.30
3	**LM_cnp**	**251.50**	**191.30**
4	RF_c	341.18	247.80
5	RF_cn	287.61	212.00
6	**RF_cnp**	**222.40**	**167.80**

Table 5.1: Table of mean RMSE and mean CRPS for evaluating model performance of the linear regression and random forest models. The two models written in bold are the models with lowest RMSE and/or CRPS score within the two different model frameworks.

In this section we evaluate the predictive performance of both the linear models and the random forest models. For evaluating their predictive performance a LOOCV was conducted, as described in section 3.6.

From **tab. 5.1** we observe that the models containing precipitation LM_cnp and RF_cnp have the highest predictive performance in terms of mean RMSE ($\overline{RMSE}$) and mean CRPS ($\overline{CRPS}$). Further we observe that the $\overline{RMSE}$ and $\overline{CRPS}$ is larger for both the linear models and random forest model with the explanatory variable/feature avg_5 (LM_cn and RF_cn) than it is for the models that only contain catchment characteristics (LM_c and RF_c). This tell us that it is not sufficient to only include catchments characteristics when modelling runoff.

The $\overline{RMSE}$ and $\overline{CRPS}$ scores are plotted against the coverage probability of a 95% prediction interval in **fig. 5.7**. This illustrates that all models have a satisfying coverage probability, as 0.95 is the best coverage probability. We further observe that the two models containing precipitation (LM_cnp and RF_cnp) have both the highest coverage probability and $\overline{RMSE}$ and $\overline{CRPS}$ score, which indicates that the models are both sharp and accurate.

From the plots with $\overline{RMSE}$ and $\overline{CRPS}$ plotted against 95% coverage probability in **fig. 5.7** we notice that the accuracy of the LM_c is better than the accuracy of LM_cn, RF_c and also that RF_cn it is less sharp as it has a higher $\overline{RMSE}$ and $\overline{CRPS}$ score. RF_c does not perform much better than the LM_c in therms of accuracy, but it has a smaller coverage probability. For the two models LM_cn and RF_cn we have that the LM_cn is better in terms of higher coverage probability and lower $\overline{RMSE}$ and $\overline{CRPS}$ score.

If we further have a look at the boxplots with RMSE and CRPS in **fig. 5.8**, we observe that all models have a long upper tail, and although the models containing precipitation (LM_cnp and RF_cnp) have smaller $\overline{RMSE}$ and $\overline{CRPS}$ scores, their upper tails are just as long as for the other models. We also notice that RF_cnp have one observation that is much larger than all other, that could be a possible outlier.

The large upper tails of RMSE and CRPS score for our models, and the high accuracy and low precision of our linear model LM_c, indicates that our models perform better for some predictions than other. By evaluating the residuals of our linear models plotted in **fig. 5.9a** we observe that our linear models does not have a constant variance σ^2 across its residuals, which indicate heteroscedasticity. The common remedy to heteroscedasticity would be to use a log-transformation. This was conducted for our linear models by log-transforming observed runoff. The results showed us that the evaluation metric RMSE performed well, but according to CRPS they performed very bad compared to all other models in this thesis. Due to readability of this result chapter we have chosen to not present the log-transformed models here. They can be found in the appendix G, where we demonstrate that the log-transformation fails due to large standard deviation of the posterior distribution.

The residuals of our random forest models plotted in **fig. 5.9b** seems be approximately random.

5.4 Main learning's

From our initial models (linear models and random forest models) the goal have been to understand what catchment characteristics are most influential for modelling runoff. We also wanted to explore how the explanatory variables/features avg_5 and precipitation influence the models and their predictive performance.

From both linear models and random models we found that the catchment characteristics are not sufficient for modeling runoff. For the linear models we observed that some of the catchment characteristics indicated predictive power when avg_5 and precipitation are not included as explanatory variables. While we observed that gradient basin indicated greater influence on our random forest models than any other feature.

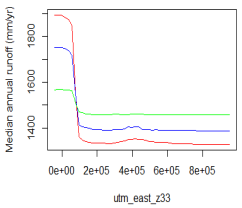
Further we have observed that avg_5 have large influence on our models compared to the catchment characteristics, but when precipitation are included into the models, the influence of avg_5 is reduced. This decreased influence of avg_5 on our models, illustrates what we already know, which is that runoff are first and foremost driven by precipitation. The decrease in influence of avg_5 on our model also illustrates how precipitation account for much of the spatial dependency between our observation that are described by avg_5.

Precipitation is obviously important for modeling runoff, but from the marginal effect of precipitation on predicted runoff illustrated in **fig. 5.5** we find that it does not seem to account for local variations in the same manner as avg_5 and gradient basin. This tells us that it is sufficient to include more explanatory variables than just precipitation into the spatial models.

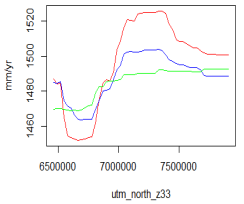
There are three catchment characteristics that describe the gradient of a catchment, gradient basin, gradient river and gradient 1085, all of which influence the linear models in terms of p-value and correlation coefficient. Gradient basin and gradient river also in-

dicated some influence on the predictive performance of our random forest models. From the explanatory analysis in chapter 2 we observed substantial correlation between the characteristics and it is thus not sufficient to use all three for our spatial models. As gradient basin is the characteristic that have largest influence on our models among the three, it is preferred for the spatial models.

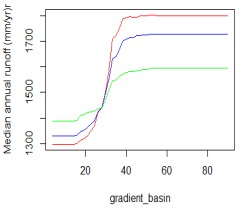
Spatial dependency seem to be the driving force of all our models. We observe that the residuals of the linear models does not fulfill the required model assumptions. The lack of homogeneity in our residuals indicate that a more direct way of modelling runoff could be sufficient.



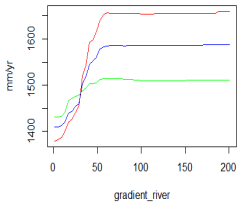
(a) UTM east coordinates.



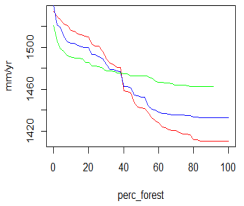
(b) UTM north coordinates.



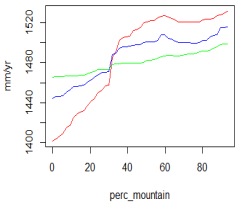
(c) Gradient basin.



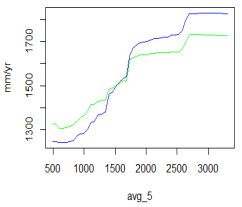
(d) Gradient river.



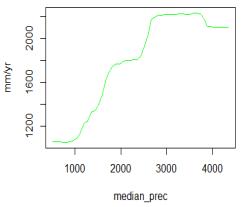
(e) Percentage of forest.



(f) Percentage of mountain.

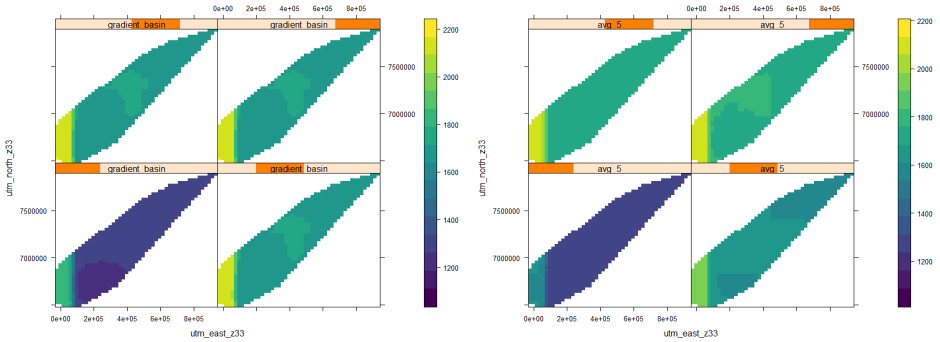


(g) Spatial dependency.



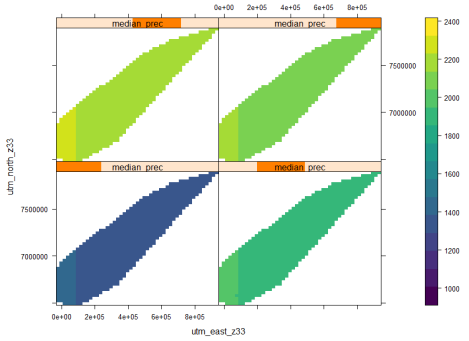
(h) Precipitation.

Figure 5.5: Partial dependency plots (pdp) of the most influential features in our random forest models. The red line belongs to marginal effect of the feature in RF_c, the blue belongs to RF_cn and the green belongs to RF_cnp.



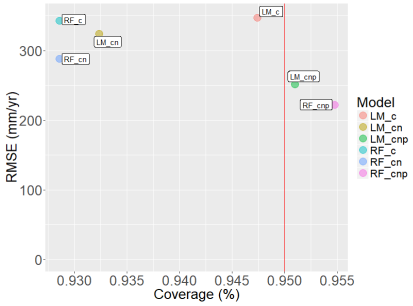
(a) Partial dependency with three features for RF_c.

(b) Partial dependency with three features for RF_cn.

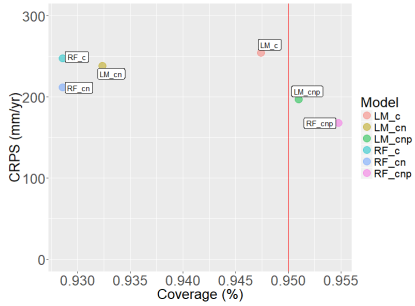


(c) Partial dependency with three features for RF_cnp.

Figure 5.6: Partial dependency in the trellis display where three features from the random forest models are displayed. On the y-axis we have the UTM north coordinates, on the x-axis we have the UTM east coordinates, and the most important features from **fig. 5.4** defines the four panels. The legend represents predicted median annual runoff (mm/yr).

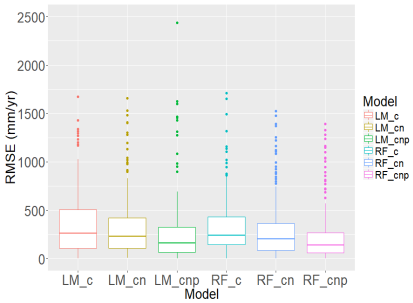


(a) |RMSE vs. 95% coverage probability.

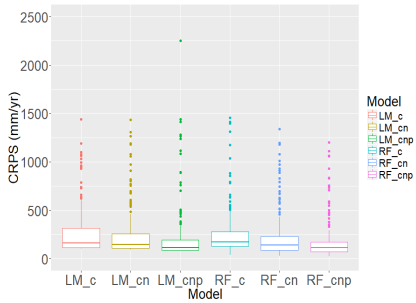


(b) |CRPS vs. 95% coverage probability.

Figure 5.7: mean RMSE and mean CRPS plotted against the 95% coverage probability for the linear models and the random forest models.

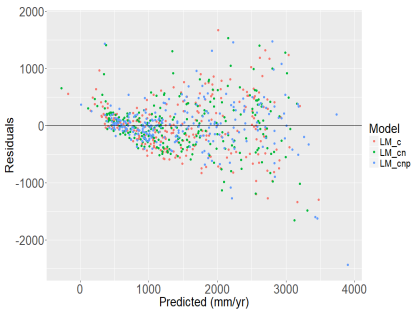


(a) RMSE score.

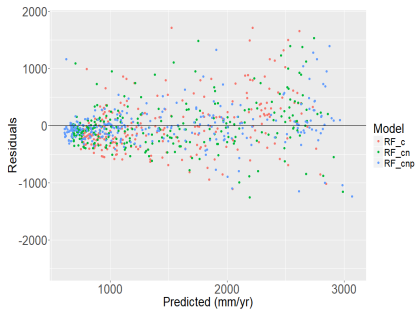


(b) CRPS score.

Figure 5.8: RMSE and CRPS score plotted as boxplots for the linear models and the random forest models.



(a) Linear models.



(b) Random forest models.

Figure 5.9: Residuals plotted against predicted runoff for the linear models and the random forest models.

Chapter 6

Results spatial models

In the following chapter we present the results of the spatial models constructed in chapter 4.3. Based on the main learning's of chapter 5.4, it seems sufficient to use a LGM allowing us to account for spatial dependence through a spatial random effect, we also include the catchment characteristics gradient basin as it seem to account for some local variations that precipitation is not able to account for.

The spatial models have been used for modeling point predictions of runoff, which we now evaluate in terms of predictive performance. The first model SP_r only uses a GRF to model runoff, from this we want to decide on whether it is sufficient to include observations of gradient basin and precipitation into the LGM. We also want to explore how well a LGM with a spatially varying coefficient for observed precipitation perform for prediction of runoff.

In this chapter, we first present the predictive performance of the four models and then we compare them. We also investigate how the models perform for different levels of runoff, and locations of catchment. Further we investigate the explanatory variables, the SPDE parameters and the GRFs to understand how the different parameters behaves for the fitted models.

6.1 Predictive performance

We now evaluate the predictive performance of the four different spatial models. The spatial models are SP_r which only includes a GRF, SP_rb with a GRF and gradient basin, SP_rbp with a GRF, gradient basin and precipitation. The forth model SP_rbpc is constructed with a GRF, gradient basin and precipitation with a spatially varying coefficient approach.

Tab. 6.1 gives the predictive performance of our spatial models in terms of $\overline{RMSE}$ and $\overline{CRPS}$ scores and also with the percentage of true runoff contained in the 95%, 65% and 45% posterior prediction intervals. Comparing SP_r and SP_rb we find a small improvement in the $\overline{RMSE}$ and $\overline{CRPS}$ score as we introduce gradient basin (SP_rb) as a covariate in the pure spatial model (SP_r). The spatial model improves further if we let the

	Model	mean RMSE	mean CRPS	Coverage95	Coverage65	Coverage45
1	SP_r	289.48	227.65	0.82	0.65	0.42
2	SP_rb	273.71	217.50	0.73	0.54	0.31
3	SP_rbp	234.68	188.42	0.76	0.46	0.23
4	SP_rbpc	**218.28**	**163.42**	0.92	0.71	0.44

Table 6.1: Table with the evaluation metrics mean RMSE and mean CRPS for the LOOCV posterior predictions for the spatial models, where SP_r only contains a GRF, SP_rb contains a GRF and gradient basin, SP_rbp contains a GRF, gradient basin and precipitation and SP_rbpc contains a GRF, gradient basin and precipitation with a varying coefficient approach. The coverage probability of the 95%, 65% and 45% posterior prediction interval is also included.

observations of precipitation (SP_rbp) be included as well. The model where the coefficient of precipitation is allowed to vary spatially (SP_rbpc) shows the best results in terms of $\overline{RMSE}$ and $\overline{CRPS}$ scores.

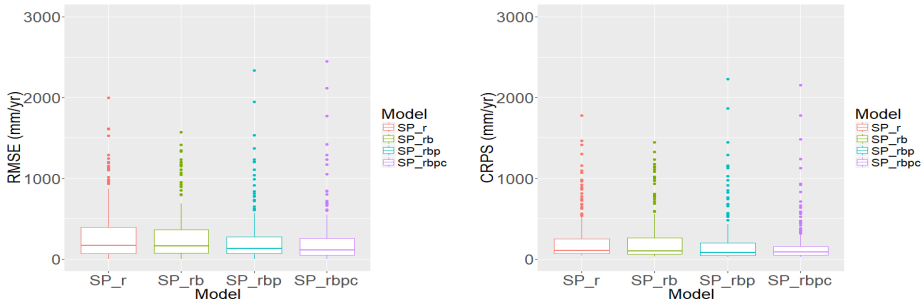
Fig. 6.1 illustrates the result of **tab. 6.1** that gives us predictive performance of our spatial models. Although the two models containing covariates (SP_rb and SP_rbp) have a lower $\overline{RMSE}$ and $\overline{CRPS}$ score than the SP_r model (which only contains a GRF), they perform worse in terms of coverage probability. Allowing us to believe that when we do not account for spatial dependency between observations in the covariates, it decreases the models performance. This is confirmed if we compare the two models SP_rbp and SP_rbpc. As SP_rbpc lets the coefficient of observed precipitation vary across our spatial domain, the coverage probability for the increases with approximately 17%, from 75% to 92%.

(a) RMSE vs. coverage **(b)** CRPS vs. coverage

Figure 6.1: The evaluation metrics mean RMSE and mean CRPS plotted against the coverage percentage of a 95% posterior prediction interval for the LOOCV predictions from the spatial models. The red line at 0.95 marks the best coverage probability.

While the $\overline{RMSE}$ and $\overline{CRPS}$ scores of SP_rbp and SP_rbpc are lower than for SP_r and SP_rb we can see from the boxplot in **fig. 6.2** that the largest RMSE and CRPS scores of SP_rbp and SP_rbpc are larger than the largest RMSE and CRPS scores for the two other models. From the boxplots with the RMSE and CRPS model it is seen that all models have some large RMSE and CRPS scores compared to the 50% quantile. If we look at the maps

plotted in **fig. F.1** and **fig. F.2** we can visualise the locations with the largest RMSE and CRPS scores. From the map we observe that the catchments with large RMSE and CRPS scores are located in coastal areas, and coastal catchments are often located in wet areas. Some of the catchments with large RMSE and CRPS scores are also located in areas containing glaciers, which in periods with warm temperatures contributes to runoff. This indicates our models perform worse for areas with more runoff than most catchments. From the map in **fig. F.1** we also observe that one catchment which is located in the middle of Norway, have much higher RMSE and CRPS score than the surrounding catchments. As this catchment deviates from all other nearby catchments it indicates a possible outlier, and should be further explored.



(a) RMSE values.

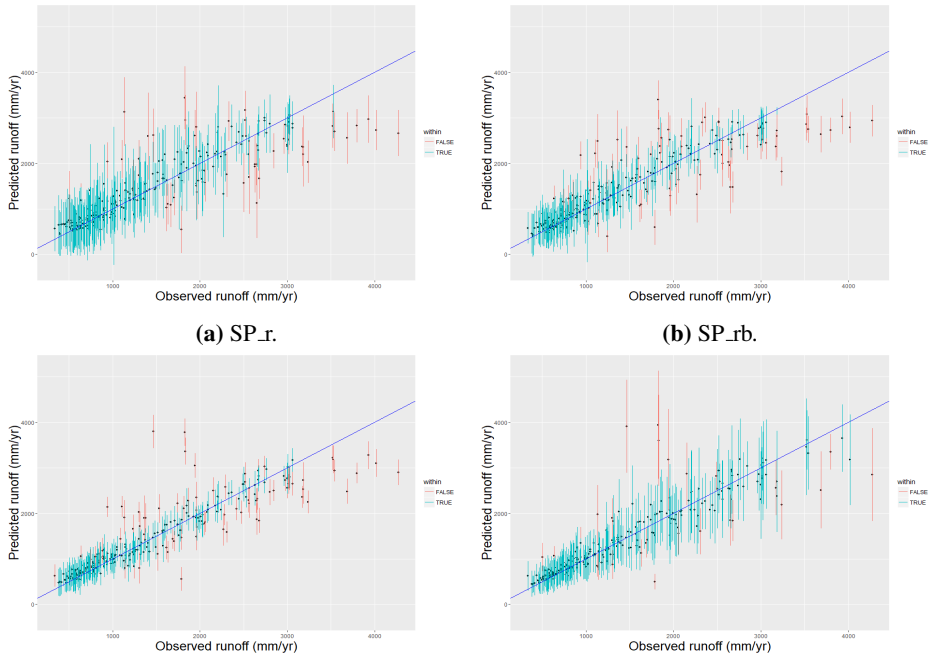
(b) CRPS values.

Figure 6.2: Boxplot of the RMSE and CRPS scores from the LOOCV predictions for the spatial models.

In **fig. 6.3** we have plots showing predicted versus observed runoff, and for each prediction the corresponding 95% posterior prediction interval is plotted. The SP_r model has a much wider prediction interval than the other models. This is why SP_r has a higher coverage probability than the two models SP_rb and SP_rbp. SP_rb does not have any higher coverage probability than SP_rbp have, even though the prediction intervals are much larger. SP_rbp have narrow prediction intervals and thus also a low coverage probability. SP_rbpc has increasing prediction intervals with increasing predicted runoff, and is thus able to cover true runoff within most predictions.

From **fig. 6.3** with plots showing predicted versus observed runoff we also observe that the two models SP_r and SP_rb have increasing deviation from observed (true) runoff as observed runoff increases. We also observe that the same occurs for SP_rbp, but the deviation from true runoff is not as large. For SP_rbpc there is also an increasing deviation from true runoff, but that most prediction intervals cover the true runoff observation. What we notice for the two models containing precipitation (SP_rbp and SP_rbpc) is that they contain two observations with much larger deviations from true runoff than for any other similar observed value, which correspond with the two large RMSE and CRPS values seen in the boxplot in **fig. 6.2**.

The increasing deviation from observed (true) runoff as observed runoff increases seen in **fig. 6.3**, indicate that the residuals are not homoscedastic, and that there are some effect in our observations of runoff which our models are not able to account for.



(a) SP_r.

(b) SP_rb.

(c) SP_rbp.

(d) SP_rbpc.

Figure 6.3: Predicted vs. observed runoff with the corresponding 95% posterior prediction interval. The colour of the prediction interval is coloured blue if the observed runoff is located within the prediction interval, and red otherwise.

The **fig. 6.4** illustrates how the residuals increase with increasing posterior predicted runoff. We also observe that the larges posterior predicted runoff values have much more negative residuals than any other posterior predicted runoff. The largest negative residuals belongs to SP_rbp and SP_rbpc, and belong to the same two catchments that we observed as large errors in the plot with posterior predicted runoff versus observed runoff in **fig. 6.3c** and in the boxplot with RMSE and CRPS in**fig. 6.3c**.

For eliminating the observed heterogeneity seen in the residual plot in **fig. 6.4**, a log-transformation of observed runoff was conducted for our spatial models. The log-transformation was not a success, according to the CRPS. For readability of the this chapter we have left out the results of the log-transformed models. These models and an exploration of their predictive performance can bee found in the appendix chapter G.

6.2 Posterior marginal distribution of the coefficients and SPDE parameters

There are some differences in the predictive performance of our spatial models. With an analysis of the posterior marginal distribution we gain knowledge about how the differ-

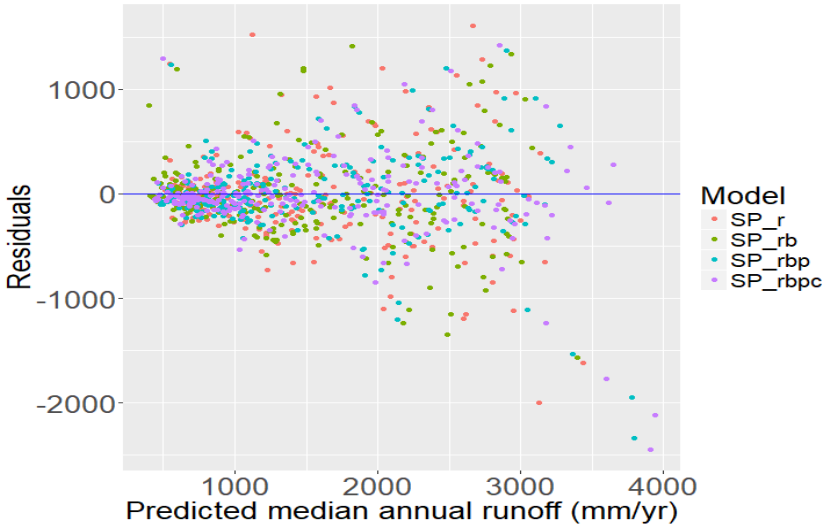


Figure 6.4: Residuals plotted against posterior predicted runoff for the four spatial models.

ent parameters behave within the models. This enable us to understand how the different terms in the models influence the posterior predictive performance. When we explore the posterior marginal distribution we explore the uncertainty and the influence the parameter have on our models, this is done by evaluating density plots. The marginal distributions are Gaussian, thus the density plots are symmetrical and centered around its mean, where the width of the density plots illustrates the variances. From the variance we obtain how uncertain the posterior marginal distribution of a coefficient is, while the mean value determines the influence a coefficient have on the model.

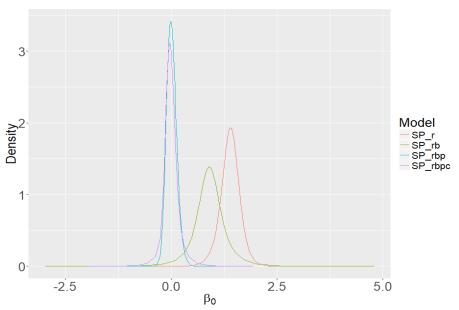
In this section we first present an evaluation of the posterior marginal distribution for the intercept, and the two coefficients of the explanatory variables gradient basin g_i and precipitation p_i. Further we evaluate the posterior marginal distribution of SPDE parameters of the GRF γ_i and the spatially random adjustment β_i.

We first explore the posterior marginal distribution of the intercept β_0 in **fig. 6.5a**, which shows that its posterior marginal distribution for the two models SP_rbp and SP_rbpc are similar. We further observe that the intercept have a larger posterior marginal mean when the observations of precipitation is not included in the two models SP_r and SP_rb. We also observe that the posterior marginal variance of SP_rbp and SP_rbpc are much smaller than for the two other models SP_r and SP_rb. This tells us that precipitation decrease the influence of the intercept and also reduce the uncertainty.

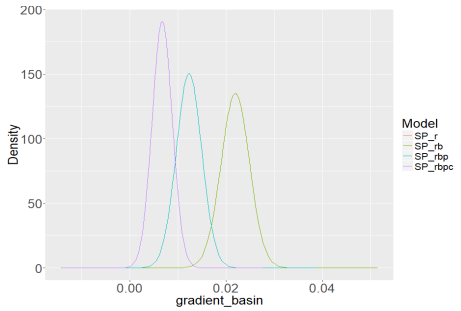
For the coefficient of gradient basin in **fig. 6.5b**, we observe that the posterior marginal distribution is more uncertain within SP_rb than it is within SP_rbp and SP_rbpc. There is also a small increase in the marginal posterior mean as we move from SP_rb to SP_rbp. This tells us that gradient basin has less effect on the model SP_rbpc than it is has for the two other models SP_rb and SP_rbp.

Further we observe from **Fig. 6.5c** that the coefficient of precipitation is more uncertain

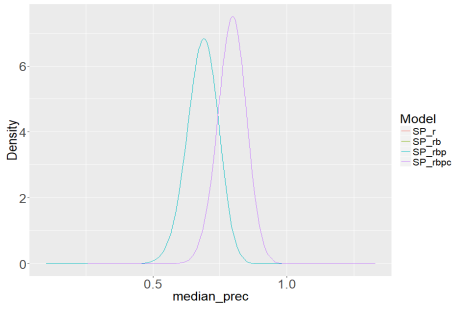
within SP_rbp than within SP_rbpc, where we allow the coefficient of precipitation to vary spatially. We also observe that the posterior marginal mean is smaller for SP_rbp than it is for SP_rbpc. The decrease in uncertainty and increase in posterior marginal mean when we allow the coefficient of precipitation to vary spatially tells us that the coefficient of precipitation in SP_rbpc have more effect on the model and is less uncertain compared to SP_rbp.



(a) Coefficient of the intercept β_0.



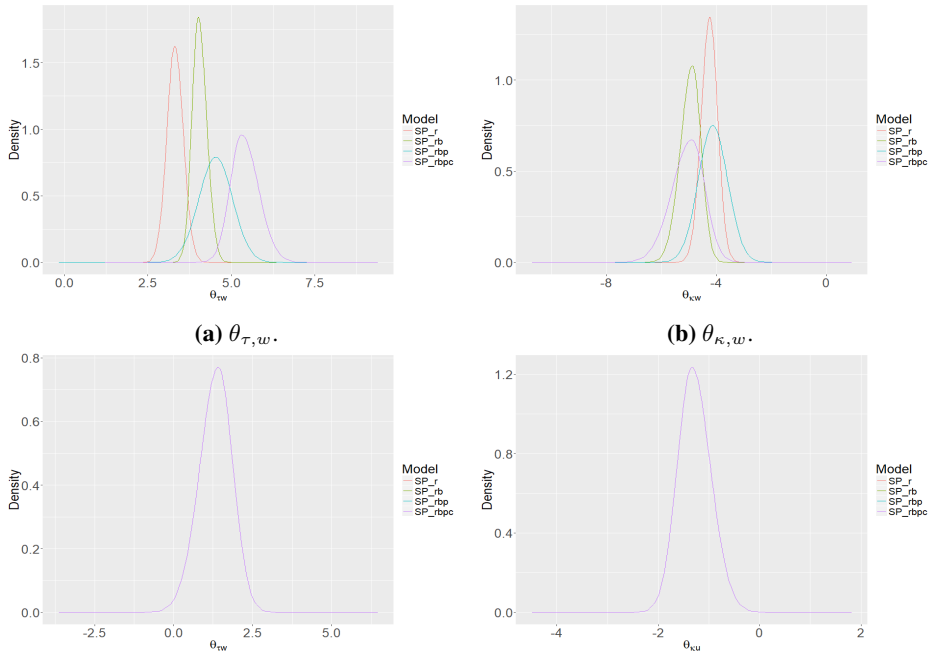
(b) Coefficient for gradient basin β_1.



(c) Coefficient for precipitation β_2.

Figure 6.5: Posterior marginal distribution of the coefficients for the intercept β_1, gradient basin β_2 and precipitation β_3 from the spatial models SP_r, SP_rb, SP_rbp and SP_rbpc.

Further we evaluate the posterior marginal distribution of the SPDE parameters $\theta_{\tau,w}$, $\theta_{\kappa,w}$ $\theta_{\tau,u}$ and $\theta_{\kappa,u}$ illustrated in **fig. 6.6**. These SPDE parameters are linked to the range and the marginal variance of the random field γ_i and the spatially varying coefficient of precipitation β_i. The posterior marginal distribution of the SPDE parameters for the spatially varying coefficients in SP_rbpc is not possible to compare with any other model, but is added to illustrate its SPDE parameters distribution. The posterior marginal distribution for the random fields SPDE parameter $\theta_{\tau,w}$ $\theta_{\kappa,w}$, shows a difference between the models. The uncertainty is much larger within SP_rbp and SP_rbcp than it is for the two other models.



(a) $\theta_{\tau,w}$. **(b)** $\theta_{\kappa,w}$.

(c) $\theta_{\tau,u}$. **(d)** $\theta_{\kappa,u}$.

Figure 6.6: Posterior distribution of the SPDE parameters for our GRF γ_i are denoted $\theta_{\tau,w}$, $\theta_{\kappa,w}$, and the SPDE parameters for our spatially random adjustment (GRF) of precipitation β_i are denoted $\theta_{\tau,u}$, $\theta_{\kappa,u}$. The SPDE parameters are linked to the range ρ and the marginal variance σ^2.

6.3 Posterior distribution of the GRFs

We now evaluate the posterior mean and standard deviation of the random field γ_i and the spatially random adjustment of precipitation β_i. To illustrate this we use maps showing their posterior mean and standard deviation within the whole study area. The two models without precipitation (SP_r and SP_rb) are very similar and we thus evaluate the posterior distribution of the random field of SP_r and look at the difference between the two models, rather than SP_rb itself. Further we evaluate the posterior distribution for the random field of SP_rbp, as it is very similar with the random field of SP_rbpc we also illustrate the difference between the two rather than SP_rbpc alone. Finally, we evaluate the spatially random adjustment of precipitation β_i.

The posterior mean of the random field for SP_r is illustrated in **fig. 6.7a**, showing that the the random field has a positive posterior mean along the coast, while the posterior mean is negative for the interior of Norway. We also observe that the posterior mean is not as high for all coastal areas of Norway, the interior also have some local variations. A positive posterior mean illustrates that the effect of the random field is large in coastal areas, and that the random field account for a negative effect in the model for the interior.

Looking at the difference between the two models SP_r and SP_rb in **fig. 6.7d**, we see

that SP_r has a larger posterior mean for the coastal areas. For the interior of Norway the random field has a smaller posterior mean for SP_r than SP_rb. This tells us that random field have a larger effect on the posterior predicted runoff for SP_r than SP_rb, while the posterior predicted runoff in the interior of Norway is less influenced by the random field of SP_r than SP_rb. The large posterior predicted mean of the random field SP_r in the coastal areas is as expected, as the random field is the only explanatory variable in SP_r.

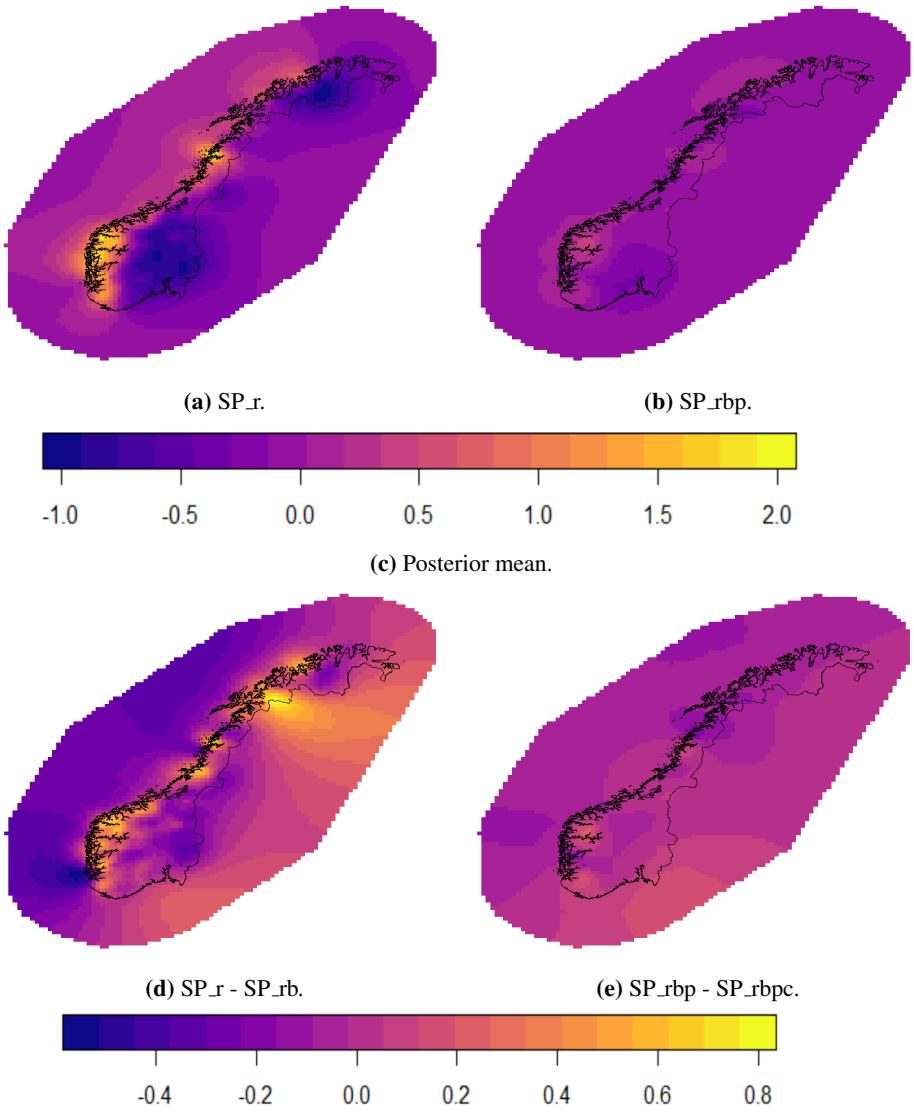
If we further evaluate the posterior standard deviation of SP_r in **fig. 6.8a**, it seems like there is an approximately constant posterior standard deviation within the whole study area, with some increase in standard deviation as we move further north in Norway. The difference in posterior predicted standard deviation of the two random field, seen in **fig. 6.8d** tells us that within the whole study area the uncertainty of SP_r is larger than within SP_rb. This reflects the decreased coverage probability of SP_rb compared to SP_r illustrated in **tab. 6.1**.

The posterior mean of the random field for SP_rbp can be seen in **fig. 6.7b**. Here we see that the posterior mean is negative within most of our study domain, which tell us that the random field of the SP_rbp has a negative effect on the posterior predicted runoff and thus punish the explanatory variables gradient basin and precipitation.

Looking at the difference between SP_rbp and SP_rbpc in **fig. 6.7e** we see some differences between the two models in the coastal areas and in the mountainous area in the interior of Norway. Illustrating that random field of SP_rbp has less effect on the posterior predicted runoff in the interior mountainous areas, and that the random field of SP_rbp account for more of the posterior predictive runoff in the coastal areas of Norway.

The posterior standard deviation of SP_rbp illustrated in **fig. 6.8b** shows that the uncertainty of the random field is constant for the whole study area, while the difference between the random fields of SP_rbp and SP_rbpc illustrated in **fig. 6.8e** shows that the posterior mean of SP_rbpc is more varying across Norway. This reflects the improved predictive performance and high coverage probability of SP_rbpc compared to SP_rbp that was shown in **tab. 6.1**.

Further we can observe the posterior mean and standard deviation of the spatially random adjustment β_i in **fig. 6.9**. Here it can bee seen that most of our study domain has a constant posterior mean of 0, while there are some local variations. For coastal areas are the effects of spatially random adjustment positive, while it is negative for most of the interior of Norway. We observe a strong positive random adjustment in the interior of Norway which belongs to the two posterior predictive runoff observations, seen in e.g. the boxplot in **fig. 6.2**. From the standard deviation of our spatially random adjustment we observe that the the uncertainty is larger in coastal areas where runoff is larger, these local variations in standard deviation of our spatially random effect illustrates why the model has the best predictive performance (see **tab. 6.1**), which increase our model performance in terms of $\overline{CRPS}$ score.



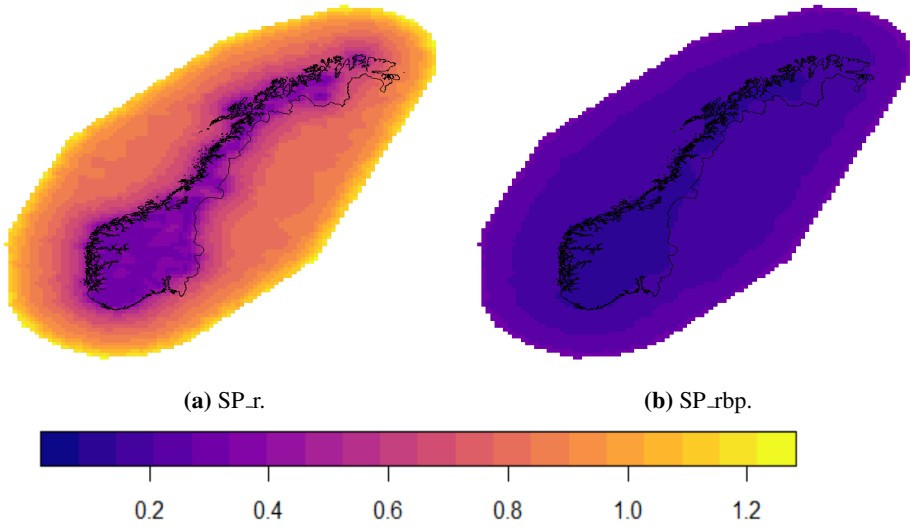
(a) SP_r. **(b)** SP_rbp.

(c) Posterior mean.

(d) SP_r - SP_rb. **(e)** SP_rbp - SP_rbpc.

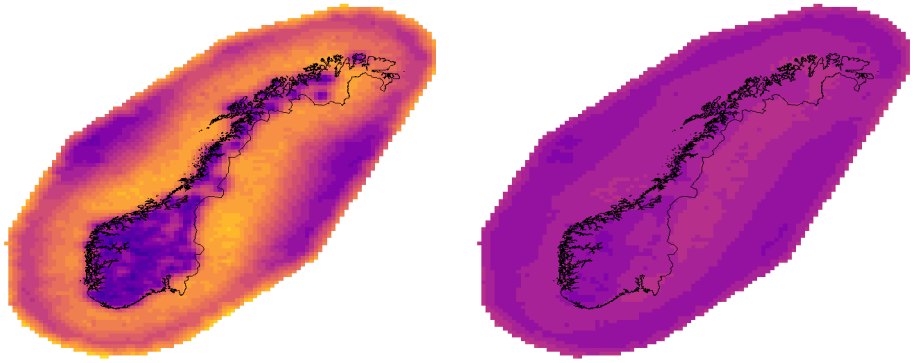
(f) Posterior mean diff.

Figure 6.7: Posterior mean of the random filed γ_i within Norway for the spatial models. (a) and (b) are the posterior mean for SP_r and SP_rbp respectively, and (d) and (e) are the difference between posterior mean of SP_r and SP_rb, and of SP_rbp and SP_rbpc respectively. Obs. the range of the two legends are different.



(a) SP_r. **(b)** SP_rbp.

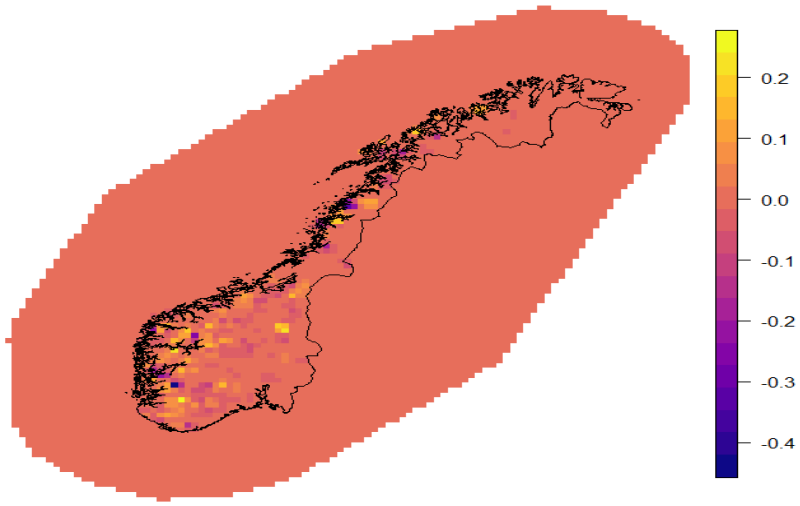
(c) Posterior standard deviation.



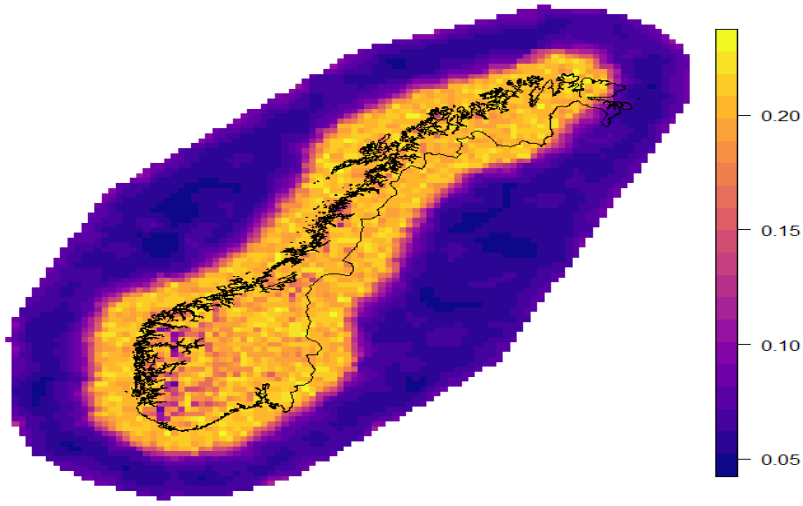
(d) SP_r - SP_rb. **(e)** SP_rbp - SP_rbpc.

(f) Posterior standard deviation diff.

Figure 6.8: Posterior standard deviation of the random field $gamma_i$ within Norway for the spatial models. (a) and (b) are the posterior standard deviation for SP_r and SP_rb respectively, and (d) and (e) are the difference between posterior standard deviation of SP_r and SP_rb, and of SP_rbp and SP_rbpc respectively. Obs. the range of the two legends are different.



(a) Posterior mean.



(b) Posterior standard deviation.

Figure 6.9: Posterior mean and standard deviation of the spatially random adjustment β_i of SP_rbpc model. Obs. the range of the two legends are different.

Chapter 7

Discussion

In this thesis we have explored different statistical models to construct runoff maps in Norway. The analysis were done in three steps. First we did explanatory analysis of catchment characteristics, runoff and precipitation data. Next, we explored two classes of initial models, linear models and random forest. From these analysis we concluded to proceed to spatial models within the Bayesian framework suited for fast computations. Within the Bayesian framework we built LGMs where the SPDE approach allow us to reduce the computational cost.

Our thesis is motivated by the runoff maps NVE creates of Norway showing mean annual runoff for the past 30 years. Our goal was to find a model able to predict runoff efficiently and with high accuracy. We explored the predictive performance by conducting a LOOCV where we used the two evaluation metrics RMSE and CRPS. While RMSE only explore the precision of our models, CRPS allow us to also explore the accuracy as it accounts for the whole posterior predictive distribution.

Through the exploration of predictive performance of our models we have found that it is not sufficient to only have a random field for modelling runoff, neither is gradient basin as an explanatory variable. The predictive performance showed large improvements as we introduced precipitation as an explanatory variable.

When we introduced precipitation into our models we observed that the intercept approached zero, which illustrates how precipitation is the main force for runoff. We also observed that gradient basin become less important as precipitation was introduced in the model. The importance of gradient basin was even less important when we introduced a spatially random adjustment of precipitation into the model. The spatially random adjustment was also able to scale the uncertainty of the model with increase observations of precipitation, such that the model become heteroscedastic.

The models where we did not allow our explanatory variable to have a spatially varying coefficient, was not able to model the heteroscedasticity in our data. And the log-transformed models was not able to improve the predictive performance of these models in terms of CRPS. A remedy to the heteroscedasticity in the models without a spatially varying coefficient could have been to allow the random error to have precision scaled

with the observations of runoff, as done in the work of Roksvåg et al. (2019) and Ingebrigtsen et al. (2015).

The model containing gradient basin only, could possibly be improved if we allowed its coefficient to vary spatially. As this would allow gradient basin to have a varying effect on the predicted runoff for different catchment locations in Norway. It would also be interesting to see if we were able to increase the predictive performance of our models if we scaled the precision of the random error.

Bibliography

Beldring, S., Roald, L. A., Voksø, A., 02 2002. Avrenningskart for norge årsmiddelverdier for avrenning 1961-1990. Tech. Rep. 2, Norges vassdrags- og energidirektorat.

Bivand, R., Keitt, T., Rowlingson, B., 2016. rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 1.1-10.
URL https://CRAN.R-project.org/package=rgdal

Bivand, R. S., Pebesma, E., Gómez-Rubio, V., 2013. Applied Spatial Data Analysis with R, second edition. Springer New York.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., Savenije, H., 2013. Runoff Prediction in Ungauged Basins. Cambridge University Press.

Breiman, L., 1996. Bagging predictors. Kluwer Academic Publishers 24, 123–140.

Breiman, L., 2001. Random forests. Kluwer Academic Publishers 45, 5–32.

Cheng, J., Xie, Y., 2016. leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library. R package version 1.0.1.
URL https://CRAN.R-project.org/package=leaflet

Cressie, N., 1993. Statistics for Spatial Data. John Wiley & Sons, Incorporated.

Devore, J., Berk, K., 2012. Modern Mathematical Statistics with Applications. Springer New York.

E. J, P., Bivand, R. S., 2005. Classes and methods for spatial data in R. Last accessed 22. May 2019.
URL https://cran.r-project.org/doc/Rnews/Rnews_2005-2.pdf

Fahrmeir, L., Kneib, T., Lang, S., Marx, B., 2013. Regression Models, Methods and Applications. Springer Heidelberg New York Dordrecht London.

Friedman, J. H., 2001. Greedy function approximation: A gradient boosting machine. The Annals of Statistics 29 (5), 1189–1232.

Gelfand, A. E., Diggle, P., Guttorp, P., Fuentes, M., 2010. Handbook of spatial statistics. Boca Raton, Fla.

Gelfand, A. E., Kim, H.-J., Sirmans, C. F., Banerjee, S., 2003. Spatial modeling with spatially varying coefficient processes. Journal of the American Statistical Association 98 (462), 387–396.

Gneiting, T., Raftery, A. E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102 (477), 359–378.

Greenwell, B., 2016. pdp: Partial Dependence Plots. R package version 0.1.0.
URL https://CRAN.R-project.org/package=pdp

Hamner, B., 2012. Metrics: Evaluation metrics for machine learning. R package version 0.1.1.
URL https://CRAN.R-project.org/package=Metrics

He, Y., Bárdossy, A., Zehe, E., 2011. A review of regionalisation for continuous streamflow simulation. Hydrology and Earth System Sciences 15 (11), 35393553.

Ingebrigtsen, R., Lindgren, F., Steinsland, I., Martino, S., 2015. Estimation of a non-stationary model for annual precipitation in southern norway using replicates of the spatial field. Spatial Statistics 14, 338–364.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning: with Applications in R. Springer New York.

Jordan, A., Krueger, F., Lerch, S., 2016. scoringRules: Scoring Rules for Parametric and Simulated Distribution Forecasts. R package version 0.9.1.
URL https://CRAN.R-project.org/package=scoringRules

Li, B., Yang, G., Wan, R., Dai, X., Zhang, Y., 2016. Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China. Hydrology Research 47 (S1), 69–83.

Liaw, A., Wiener, M., 2002. Classification and regression by randomforest. R News 2 (3), 18–22.
URL http://CRAN.R-project.org/doc/Rnews/

Lindgren, F., Rue, H., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73 (4), 423–498.

Lindgren, F., Rue, H., 2015. Bayesian spatial modelling with R-INLA. Journal of Statistical Software 63 (1), 1–25.

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., Andersen, J., 2018. senorge2 daily precipitation, an observational gridded dataset over norway from 1957 to the present day. Earth Syst. Sci. Data 10, 235–249.

Parajka, J., Merz, R., Blöschl, G., 2005. A comparison of regionalisation methods for catchment model parameters. Hydrology and Earth System Sciences 9 (1-2), 157–171.

Parajka, J., Viglione, A., Rogger, M., Salinas, J., Sivapalan, M., Blöschl, G., 2013. Comparative assessment of predictions in ungauged basins - part 1: Runoff-hydrograph studies. Hydrology and Earth System Sciences 17 (5), 1783–1795.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL http://www.R-project.org/

R Core Team, 2016. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/

Robert, C., Casella, G., 2004. Monte Carlo statistical methods 2nd ed. Springer New York.

Roksvåg, T., Steinsland, I., Engeland, K., 2019. A knowledge based spatial model for utilising point and nested areal observations: A case study of annual runoff predictions in the Voss area. arXiv:1904.02519.

Rue, H., Held, L., 2005. Gaussian Markov random fields: theory and applications. Chapman and Hall/CRC.

Rue, H., Martino, S., Chopin, N., 2009. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. Journal of the Royal Statistical Society: Series B (Statistical Methodology 71 (2), 319–392.

Sauquet, E., Gottschalk, L., Leblois, E., 2000. Mapping average annual runo: A hierarchical approach applying a stochastic interpolation scheme. Hydrological Sciences Journal 45 (6), 799–815.

Skøien, J. O., Merz, R., Blöschl, G., 2006. Top-kriging - geostatistics on stream networks. Hydrology and Earth System Sciences 10 (2), 277–287.

Statkraft, 2016. Vannkraft. Last accessed 23. May 2019.
URL https://www.statkraft.no/Energikilder/Vannkraft/

Strahler, A. N., 1952. Hypsometric (area-altitude) analysis of erosional topography. Geological Society of America Bulletin 63 (11), 11177606.

Tobler, W. R., 1970. A computer movie simulating urban growth in the detroit region. Economic Geography 46, 234–240.

White, E., 2015. Predicting unimpaired flow in ungauged basins: random forests applied to california streams. Last accessed 19. December 2018.
URL https://watershed.ucdavis.edu/shed/lund/students/EllieWhiteMSthesis.pdf

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.
URL http://ggplot2.org

Wickham, H., Francois, R., Henry, L., Müller, K., 2017. dplyr: A Grammar of Data Manipulation. R package version 0.7.4.
URL https://CRAN.R-project.org/package=dplyr

Wright, S., 1921. Correlation and causation. Journal of Agricultural Research 7, 281–305.

Yang, X., Magnusson, J., Rizzi, J., Xu, C.-Y., 2018. Runoff prediction in ungauged catchments in Norway: comparison of regionalization approaches. Hydrology Research 49 (2), 487–505.

Appendix A

Suspect observations and data selection

For this thesis we have emphasized the uncertainty within our models. One part of reducing the uncertainty is to carefully choose the data, and remove outliers and suspicious observation. We will now present how we have selected our data, what data was left out and also reasoning for these choices.

A.1 Data selection

Removing catchments based on time

At NVE they create runoff maps of 30-year periods. As this thesis is based on the same data as they will use in their runoff map of Norway for the period 19912020, and also motivated by this work we have chosen to look at the 30-year period from 19872016.

From NVE we received observed runoff from 699 catchments. Each catchment has been carefully considered by NVE and they have criterias that each catchment has to fulfil in order to be used. The requirements are that they only consider catchments with data from after 1958 and they have to have at least 5 years of data. Catchments that have been affected by regulations are left out, if a catchment has been regulated in the past, but not anymore, they can use the parts unaffected.

When we choose to focus on the hydrological years 19872016, 16 of our catchments provided by NVE was removed. The catchments removed can be seen in **tab. A.1**.

Further we decided that in order to minimize uncertainty we choose to only use the years within each catchment that contained at least 99.5% of the data. If more than 99.5% of the data was missing, the year was removed.

It was also required that each catchment had to have at least 10 years of overlap of data. This reduced number of catchments considerably, 308 catchments did not meet this requirement. The 308 catchments left out of our dataset are listed in **tab. A.2**.

Removing catchments based on catchment characteristics

After removing catchments with missing observations within our time period we are left with 341 catchments. We know that some catchments also have missing catchment characteristics. The catchment characteristics with missing observations can be found in **tab. A.3**. With some simple initial analysis of our dataset, gradient basin seems too be a catchment characteristic we want to carry on with, so we only remove the catchments that are missing this catchment characteristic. We also remove the catchments that have missing observations of gradient 1085 and length of the basin. This leaves us with 268 catchments. The catchments we remove can be found in **tab. A.4** and **tab. A.5**.

Field ID	Station name
163	Fossum bru
179	Visa
215	Fundin ndf.
343	Gryta
368	Sæternbekken
505	Strøen ndf.
678	Bitdalsvatn
863	Stegemoen
878	Homstølvatn
919	Lundevatn
1460	Sørdalsvatn
1585	Elverhøy bru
1660	Gaulfoss
1699	Stokke
1873	Søndre Bjøllåvatn
2238	Galten

Table A.1: Catchments with missing observations in the hydrological time period 19872017.

Removing catchments based on strange observations

While working on the initial analysis some time was spent on analysing values of yearly runoff. In **Fig. A.1** histograms of yearly runoff for our catchments are displayed. Also a histogram comparing median runoff from the periods of 19872016 and 19611990 can be seen in **Fig. A.2**.

First investigating **Fig. A.1**, the histogram illustrating maximum yearly runoff. Catchment named, *Flostrand* (field ID 1888), has a maximum value of 6567.0 mm/yr and a minimum of 2937.1 mm/yr, it also has a high standard deviance of 789 mmyr. And another 9 catchments also have a maximum yearly runoff above 5000 mm/yr.

Looking at minimum yearly runoff in **Fig. A.1**. The larges minimum value belongs to the catchment named, *Skjerdalselv* (field ID 1450), which has a minimum of 3380 mm/yr, a maximum of 5294 mm/yr, a median of 4271 mm/yr and a standard deviance of 556

mm/yr. The smallest observation belongs to the catchment named, *Dorgefoss* (field ID 886), it has a minimum of 11 mm/yr, maximum of 286 mm/yr, median of 21 mm/yr and a standard deviation of 50.1 mm/yr.

In the 95% quantile histogram (**Fig. A.1**) we observe that the largest runoff observation, also belong to the larges values observed in the histogram with maximume runoff **Fig. A.1**. The largest runoff value belongs to the catchment *Flostrand* (field ID 1888) with a runoff of 5602 mm/yr. And the smallest observed runoff value belongs to the catchment *Dorgefoss* (field ID 886) with a value of 85 mm/yr.

In the 5% quantile histogram in (**Fig. A.1**). This also shows that the smallest value belongs to *Dorgefoss*, and the larges to *Flostrand*.

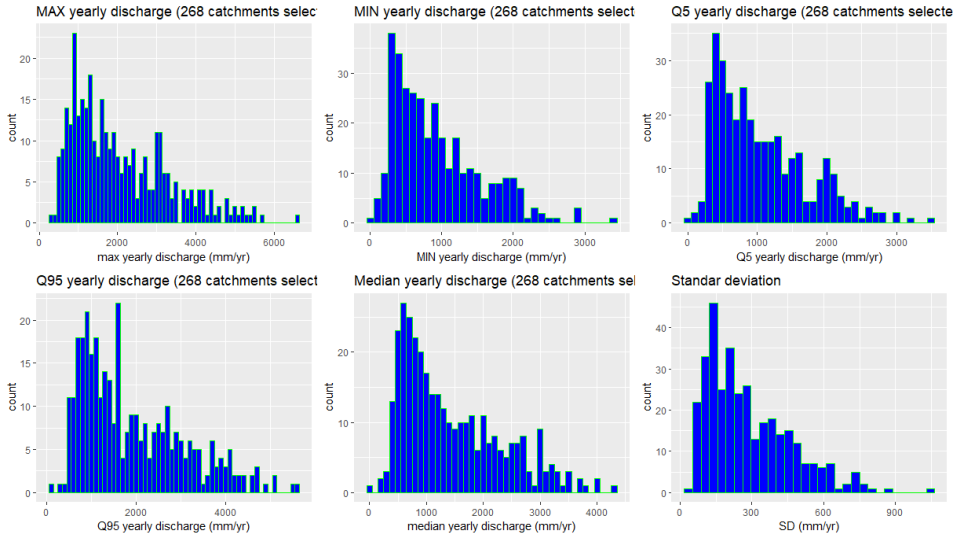


Figure A.1: Histogram of maximum runoff, minimum runoff, 5% quantile runoff, 95% quantile runoff, median runoff, and standard deviation runoff.

Fig. A.2 shows the median annual runoff from our data (red), and mean annual runoff within the time period 19611990. We have used this as an indication of what observations might be strange. By looking at the shapes of the two histograms everything seems mostly fine. When calculating the difference between the observations in median runoff for 19611990 and median runoff for 19872017, we get a mean difference of 4.3 mm/yr, median difference of -45.4 mm/yr, the maximum difference is of 2710.5 mm/yr and a minimum of -1352.7 mm/yr.

From **fig. A.2** we see that the smallest observation in red belongs to *Dorgefoss* with a value of 21.5 mm/yr. In blue this catchment has a value of 2413 mm/yr. If we now look at **tab. A.6** and **tab. A.7** where tables of the observations with larges difference is displayed, we see that the difference in runoff between the two time periods for *Dorgefoss* is of 2391 mm/yr, which is a large difference. **Tab. A.7** and **tab. A.6** shows several catchments with large difference, but most of them are either located within glaciers or in areas with large precipitation. We also notice *Valle* (field ID 801) in **tab. A.7** which has a low minimum

median annual runoff and a high difference compared with the data from 19611990.

After consulting my findings with experts at NVE we have also decided to leave *Dorgefoss* and *Valle* out of our data reducing our dataset to a number of 266 catchments.

Figure A.2: Histogram comparing median runoff in 1961-1990 period with the 1987-2017 period.

field ID	field ID	field ID	field ID	field ID	field ID	field ID
23	511	857	1124	1587	1914	2171
45	517	858	1162	1592	1917	2181
50	520	864	1164	1600	1920	2183
51	525	870	1170	1628	1922	2184
65	547	872	1179	1632	1941	2190
67	548	875	1180	1633	1942	2201
76	549	885	1193	1663	1943	2203
78	550	887	1195	1679	1944	2204
132	551	888	1212	1686	1954	2211
141	565	889	1213	1694	1959	2212
144	583	891	1225	1702	1966	2213
153	588	894	1234	1709	1967	2223
164	591	895	1242	1710	1973	2224
167	594	901	1262	1712	1975	2233
173	597	920	1265	1713	1986	2234
174	601	930	1267	1725	1990	2240
175	637	943	1288	1730	1996	2250
180	651	956	1323	1740	1997	2263
182	662	957	1338	1749	1998	2272
183	663	958	1352	1772	1999	2273
184	668	962	1395	1788	2000	2274
187	669	963	1404	1816	2003	2275
196	679	966	1407	1820	2011	2278
205	684	970	1417	1821	2016	2283
214	687	986	1420	1824	2031	2284
218	701	989	1422	1828	2041	2289
219	710	992	1424	1829	2043	2290
220	711	993	1431	1832	2053	2294
241	714	994	1432	1840	2056	2300
250	720	1010	1441	1841	2059	2309
260	723	1011	1443	1847	2063	2323
279	728	1013	1445	1855	2083	2325
283	740	1015	1474	1857	2104	2330
284	748	1041	1475	1858	2105	2336
294	749	1052	1478	1862	2112	2339
312	759	1060	1479	1868	2114	2350
317	760	1065	1487	1870	2115	2351
323	761	1072	1497	1871	2127	2359
327	762	1079	1526	1872	2135	2369
370	779	1092	1545	1889	2137	2385
378	782	1095	1575	1894	2138	2417
452	807	1101	1578	1896	2145	2431
456	832	1113	1579	1907	2152	2483
488	847	1118	1582	1913	2161	

Table A.2: Field ID for the catchments that do not have a 10 year overlap.

Catchment characteristic	number of missing data
gradient_basin	70
gradient_1085	2
lenght_km_basin	2

Table A.3: Catchment characteristics with missing data.

feltNr	gradient_basin	gradient_1085	length_km_basin
17		0.29	104.86
305		1.89	226.04
330		3.34	34.28
719	22.04		2.40
860		14.63	19.54
1053		29.43	20.99
1055		34.92	23.07
1058		55.74	
1061		83.10	
1353		28.53	14.71
1418		62.88	11.72
1470		33.84	33.46
1472		5.89	21.70
1480		51.83	21.07
1576		10.61	44.62
1577		35.15	16.72
1580		25.50	12.91
1584		59.14	8.03
1614		26.24	6.28
1664		20.03	23.20
1665		27.57	14.35
1666		21.06	35.03
1768		9.10	21.78
1782		4.38	125.36
1784		24.00	14.55
1789		12.21	34.16
1790		7.86	45.41
1796		5.99	74.68
1823		21.29	11.63
1825		14.52	8.80
1827		8.23	31.49
1861		36.80	14.83
1864		22.83	35.84
1867		26.50	11.12

Table A.4: Catchments with missing catchment characteristics, part 1.

Field ID	gradient_basin	gradient_1085	length_km_basin
1874		27.46	8.98
1876		40.31	11.10
1932		15.60	18.32
1938		25.87	29.78
1939		14.17	14.69
2022		15.30	36.44
2026		98.53	6.65
2047		21.91	7.28
2052		64.27	3.67
2064		6.47	38.87
2081		15.04	48.93
2090		11.25	41.01
2094		7.48	80.70
2095		4.90	84.36
2103		30.36	13.67
2116		31.09	11.91
2126		5.57	64.66
2168		6.36	31.62
2180		0.99	219.73
2185		1.71	50.97
2189		1.68	50.94
2194		1.00	218.32
2200		21.15	5.95
2206		1.82	89.38
2207		1.74	88.10
2214		3.51	23.36
2216		10.26	18.03
2225		8.85	28.52
2228		1.22	68.87
2230		2.35	142.63
2242		5.74	26.19
2253		4.33	372.51
2255		3.12	146.87
2261		16.67	22.63
2264	273.87		69.74
2279		4.98	39.14
2280		6.35	23.88
2281		5.12	13.21
2460		0.90	224.46

Table A.5: Catchments with missing catchment characteristics, part 2.

Field ID	Runoff 2017	Runoff 1990	Diff	MIN	MAX	SD
1440	1827.90	4538.37	2710.47	1219.76	2995.03	472.16
1138	1830.03	4313.63	2483.60	1094.88	2517.64	373.76
886	21.50	2412.74	2391.24	10.73	285.64	50.13
1893	1463.91	3500.88	2036.97	880.45	4602.84	1061.99
1076	1943.34	3470.67	1527.33	1609.89	3326.43	455.83
801	215.95	1659.83	1443.88	144.56	397.97	53.20
1800	1533.22	2608.92	1075.70	1011.44	2428.67	299.54
1799	2502.79	3484.82	982.03	1908.33	3485.67	441.88
1004	938.94	1919.00	980.06	787.95	1457.97	199.35
820	1743.78	2613.12	869.34	1206.36	2961.54	410.08
977	2655.68	3409.94	754.26	1802.04	4087.15	611.72
1007	1108.58	1671.00	562.42	881.72	1588.62	182.42
1051	2669.56	3225.26	555.70	1744.44	5480.08	778.70
1345	1297.32	1846.41	549.09	804.86	2733.72	406.51
1344	1740.13	2256.22	516.09	1482.14	2661.23	258.41

Table A.6: Highest positive difference between median annual runoff in 1961-1990 and mean annual runoff in 1987-2017.

Field ID	Runoff 2017	Runoff 1990	Diff	MIN	MAX	SD
1895	2999.52	2347.15	-652.37	2022.23	4768.15	631.88
1888	4002.52	3317.25	-685.27	2937.05	6566.95	859.01
1711	1252.85	503.00	-749.85	934.63	1611.54	219.98
1559	2680.94	1780.63	-900.31	2036.47	3457.11	435.73
2136	2650.66	1648.48	-1002.18	1962.49	3036.39	293.81
89	1789.31	436.60	-1352.71	1190.78	1981.00	218.93

Table A.7: Highest negative difference between median annual runoff in 1961-1990 and mean annual runoff in 1987-2017.

Field ID	Field ID	Field ID	Field ID	Field ID	Field ID	Field ID	Field ID
17	317	709	958	1331	1632	1876	2135
22	323	710	962	1338	1633	1887	2136
23	327	711	963	1339	1639	1888	2137
29	330	714	966	1344	1644	1889	2138
30	343	718	970	1345	1660	1893	2145
32	365	719	977	1352	1661	1894	2147
39	368	720	986	1353	1663	1895	2148
40	370	723	989	1363	1664	1896	2149
43	378	728	992	1367	1665	1907	2152
45	392	740	993	1383	1666	1913	2156
50	393	746	994	1393	1679	1914	2157
51	424	747	997	1395	1683	1917	2160
65	426	748	1003	1396	1684	1918	2161
67	445	749	1004	1401	1686	1920	2167
76	452	750	1007	1402	1689	1922	2168
78	454	751	1010	1404	1691	1932	2171
81	455	752	1011	1405	1694	1938	2177
86	456	753	1013	1406	1699	1939	2180
89	474	759	1014	1407	1701	1940	2181
95	475	760	1015	1417	1702	1941	2183
97	485	761	1032	1418	1709	1942	2184
128	488	762	1041	1420	1710	1943	2185
132	502	763	1051	1421	1711	1944	2189
141	505	767	1052	1422	1712	1954	2190
144	507	770	1053	1424	1713	1959	2194
153	511	775	1054	1425	1714	1965	2200
158	513	776	1055	1429	1715	1966	2201
159	514	777	1058	1431	1723	1967	2203
160	515	779	1060	1432	1725	1972	2204
161	516	782	1061	1433	1727	1973	2206
163	517	784	1063	1434	1729	1975	2207
164	518	801	1065	1440	1730	1986	2211
166	520	807	1072	1441	1731	1987	2212
167	524	820	1076	1443	1736	1988	2213
168	525	822	1079	1445	1739	1990	2214
169	527	831	1092	1448	1740	1996	2216
172	529	832	1094	1450	1749	1997	2219
173	537	840	1095	1460	1753	1998	2220
174	545	844	1098	1468	1762	1999	2223
175	547	846	1101	1470	1765	2000	2224
176	548	847	1109	1471	1768	2003	2225
177	549	853	1113	1472	1772	2005	2228

Table A.8: First half of field IDs for catchments received by NVE.

Field ID	Field ID	Field ID	Field ID	Field ID	Field ID	Field ID	Field ID
179	550	857	1118	1474	1782	2008	2230
180	551	858	1121	1475	1784	2011	2233
182	561	860	1124	1478	1788	2016	2234
183	565	861	1136	1479	1789	2022	2238
184	567	863	1138	1480	1790	2026	2240
186	569	864	1157	1486	1796	2031	2242
187	583	869	1162	1487	1797	2041	2250
195	586	870	1163	1497	1799	2043	2253
196	588	872	1164	1502	1800	2047	2255
201	589	875	1167	1504	1808	2048	2261
205	591	878	1170	1513	1810	2052	2263
206	593	879	1171	1526	1811	2053	2264
214	594	881	1172	1530	1816	2055	2272
215	597	885	1175	1532	1820	2056	2273
218	599	886	1179	1534	1821	2057	2274
219	601	887	1180	1535	1823	2059	2275
220	602	888	1187	1544	1824	2060	2278
231	612	889	1193	1545	1825	2063	2279
241	618	891	1195	1559	1827	2064	2280
242	635	894	1212	1560	1828	2065	2281
248	637	895	1213	1561	1829	2068	2283
250	643	899	1225	1573	1832	2077	2284
252	645	900	1226	1575	1835	2081	2289
255	651	901	1232	1576	1836	2082	2290
260	662	905	1234	1577	1840	2083	2294
263	663	908	1239	1578	1841	2090	2300
267	668	910	1242	1579	1847	2094	2309
268	669	919	1256	1580	1855	2095	2323
269	672	920	1257	1582	1857	2101	2325
270	677	929	1262	1584	1858	2103	2330
271	678	930	1265	1585	1861	2104	2336
272	679	931	1267	1587	1862	2105	2339
279	680	932	1272	1592	1864	2112	2350
283	682	935	1277	1593	1865	2113	2351
284	684	939	1288	1594	1867	2114	2359
294	687	940	1298	1600	1868	2115	2369
304	693	941	1300	1602	1870	2116	2385
305	694	943	1302	1614	1871	2121	2417
311	698	949	1323	1621	1872	2125	2431
312	701	956	1325	1628	1873	2126	2460
313	702	957	1326	1630	1874	2127	2483

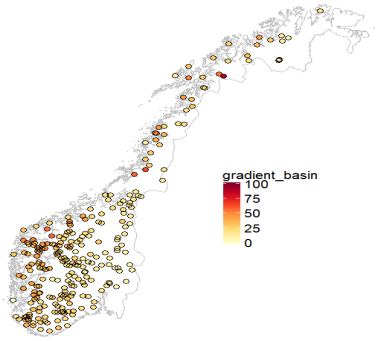
Table A.9: Second half of field IDs for catchments received by NVE.

Field ID	Field ID	Field ID	Field ID	Field ID	Field ID	Field ID
248	40	1534	1739	1799	1138	752
81	454	255	272	2060	677	1559
231	612	680	1723	502	1171	1187
424	1727	746	128	910	770	1051
97	304	507	1448	899	1363	527
693	160	2219	1167	514	1502	753
268	475	1630	1753	1175	1530	1383
2149	1715	776	2157	941	672	1226
2147	1573	861	1691	1560	22	2055
2148	635	524	949	767	201	694
593	455	1232	1513	1406	1940	643
445	1661	1239	1561	1339	698	1811
393	2113	2125	1593	747	1014	751
561	940	169	2101	931	682	1393
39	474	1121	2160	1076	1888	513
775	158	177	267	935	1440	1689
618	1468	176	2068	586	1063	1054
537	1344	2082	32	853	1987	977
569	1429	1621	567	599	905	763
545	426	718	86	2220	1277	784
1683	1256	2156	932	1865	1367	
161	159	1421	820	929	1302	
1639	1835	1109	1098	2077	900	
518	2065	1918	529	1300	1988	
311	1172	1433	29	908	1272	
1711	1401	269	702	1972	1486	
777	1701	1396	645	515	485	
831	186	166	1434	1893	1594	
252	30	1965	1808	392	1298	
1007	1736	1762	1765	1136	1450	
1004	1094	1402	1405	881	997	
95	43	1325	879	1714	1895	
602	271	846	1729	313	2008	
89	172	869	1810	1836	1032	
709	1731	1471	1644	1331	2177	
1535	263	2121	242	195	750	
1163	1532	270	589	2057	822	
1684	168	365	206	1326	2048	
844	1797	939	2005	1345	1887	
1257	1800	516	1504	1003	1157	
2167	1425	840	1602	1544	2136	

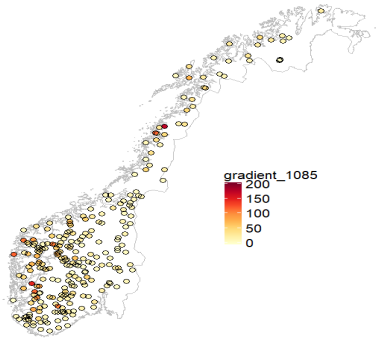
Table A.10: List of field ID for all 266 catchments used

Appendix B

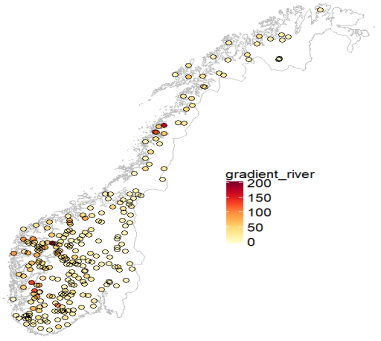
Maps of catchment characteristics



(a) Gradient basin (m/km).

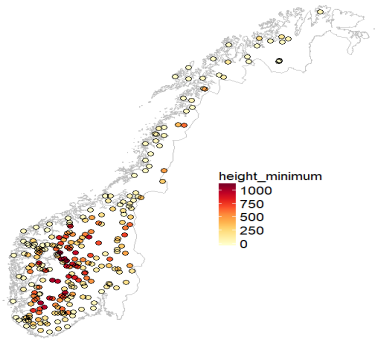


(b) Gradient 1085 (m/km).

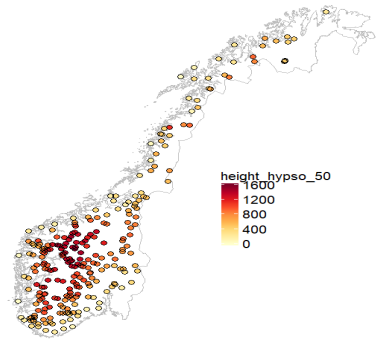


(c) Gradient river (m/km).

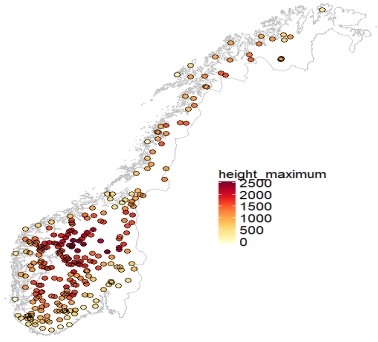
Figure B.1: Map of gradients.



(a) Height minimum (m. a.s.l).

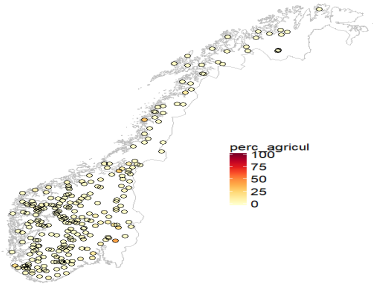


(b) Height hypso 50 (m. a.s.l).

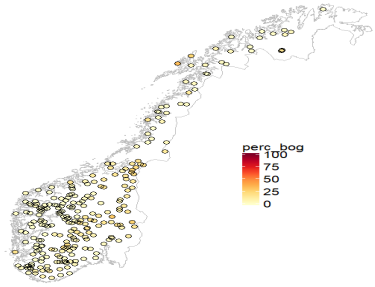


(c) Height maximum (m. a.s.l).

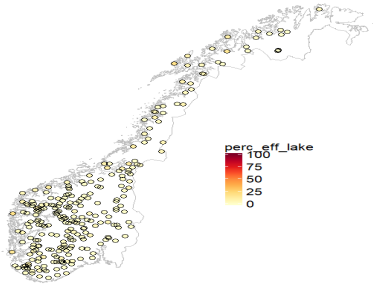
Figure B.2: Map of elevation.



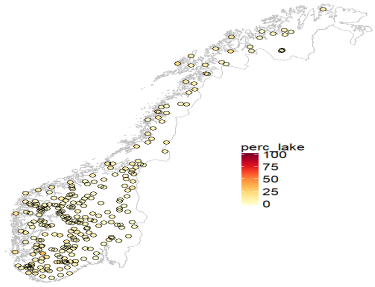
(a) Percentage agriculture.



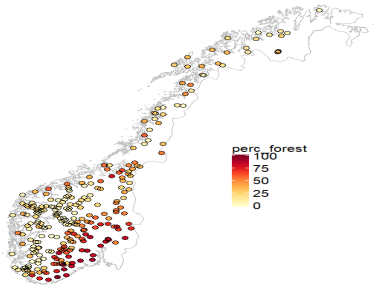
(b) Percentage bog.



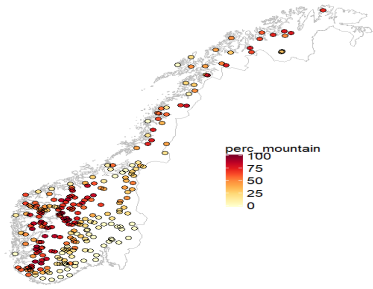
(c) Percentage effective lake.



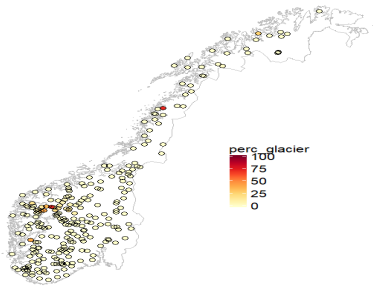
(d) Percentage lake.



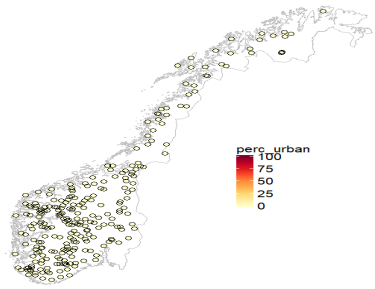
(e) Percentage forest.



(f) Percentage mountain.

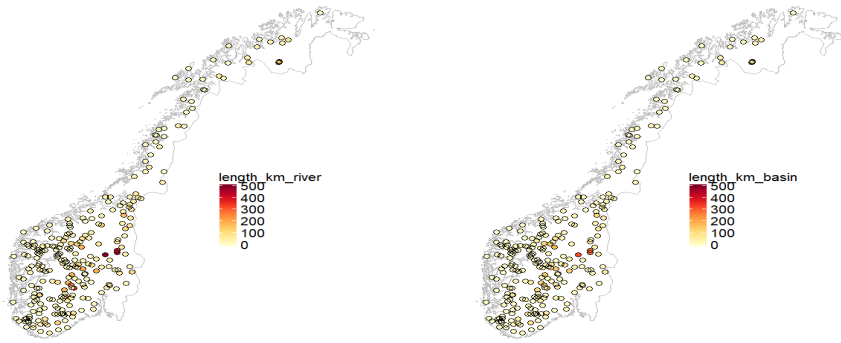


(g) Percentage glacier.



(h) Percentage urban.

Figure B.3: Map of elevation.



(a) Length river (km).

(b) Length basin (km).

Figure B.4: Map showing lengths of rivers and basins.

Appendix C

Exploration of spatial dependency between observations of yearly runoff

To explore the spatial correlation between observations of yearly runoff we use a semivariogram plot. We use this semivariogram as a tool for exploring whether there are any spatial dependency between our catchments. If there is spatial dependency, we are able to get an approximate range of how far the correlation reaches.

The variogram is defined by Cressie (1993) as

$$2\gamma(s_1 - s_2) = \text{Var}(\eta_1 - \eta_2) \tag{C.1}$$

where $2\gamma(\cdot)$ is the variogram of the random process η defined as $\{\eta(\mathbf{s}) : \mathbf{s} \in D\}$ where $\mathbf{s_i}$ are locations in $\mathbf{R}^2$ for $i = 1, .., n$.

The model illustrated below in **Fig. C.1** is a semivariogram $\gamma(\cdot)$ where a Gaussian model has been fitted. The 266 catchments have been divided into 15 groups based on distance of separation. In the plot we can see that the distance ranges from 75 km to 1000 km. By looking at the semivariogram we see that it flattens out at a range of approximately 200 km. Telling us that there probably are some correlation between catchments, but if the range between catchments becomes greater than approximately 200 km it does not seems to be any correlation.

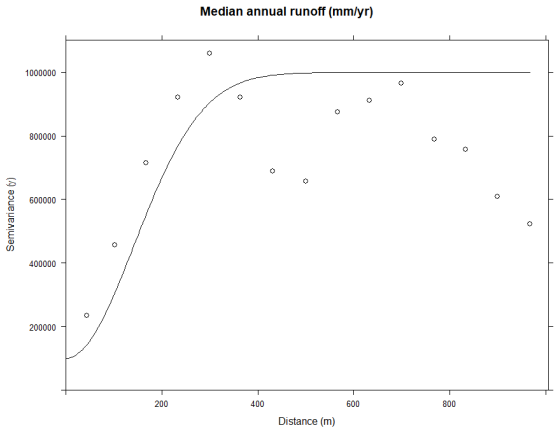


Figure C.1: Semivariance plot of the 266 catchments divided into 15 groups. The nugget is where the semivariogram model intercepts the y-axis, here it is at approximately 100 000. The range is where the model first flattens out, which here is at approximately 200 km. The sill is the value of semivariance where our model attains its range.

Appendix D

The LGM

In this chapter we build the LGM presented in sections 3.2 suited for INLA, which we use for predicting runoff $\mathbf{y}$. We first present the Bayesian hierarchical construction of the pure model SP_r, this model is the "baseline" for all our spatial models, afterwards we present an overview of the LGMs SP_rb, SP_rbp and SP_rbpc.

The observation model **eq. 4.3** has a Gaussian distributed random error with mean zero and prePIsion τ_p^{-1}, and thus the observation model is Gaussian distributed dependent on the GMRF $\mathbf{w}$, the intercept β_0 and the prePIsion τ_p^{-1}, e.g.

$$\mathbf{y}|\mathbf{w}, \beta_0, \tau_p \sim N(\eta, \tau_p^{-1}\mathbf{I}) \tag{D.1}$$

with the expected value being

$$E[\mathbf{y}|\mathbf{w}, \beta_0, \tau_p] = \eta(\mathbf{s}) \tag{D.2}$$

where $\eta(\mathbf{s})$ is the model for the unobserved process (field) as defined in **eq. 4.12**.

Further we gather the parameters of the unobserved random process $\eta(\mathbf{s})$ in a vector, $\theta_1 = [\mathbf{w}, \beta_0]$, where our spatial parameter $\mathbf{w}$ is a GMRF with the prePIsion matrix given in **eq. 3.25**.

$$\mathbf{w}|\theta_{\tau,w}, \theta_{\kappa,w} \sim N(0, \mathbf{Q}^{-1}(\theta_{\tau,w}, \theta_{\kappa,w})) \tag{D.3}$$

$$\beta_0 \sim N(\cdot, \cdot) \tag{D.4}$$

since all the parameters of the unobserved process $\eta(\mathbf{s})$ are all Gaussian they are also jointly Gaussian e.g. $\theta_{\tau,w} \sim N(\cdot, \cdot)$.

For our spatial models we have the three SPDE parameters τ_p, $\theta_{\tau,w}$ and $\theta_{\kappa,w}$ gathered in the following vector $\theta_{\mathbf{2}} = [\tau_p, \theta_{\kappa,w}, \theta_{\tau,w}]$. $\theta_{\mathbf{2}}$ defines our last level of our Bayesian hierarchical model, and the joint distribution of our model for the hyper parameters $\theta_{\mathbf{2}}$ are defined as

$$\pi(\theta_{\mathbf{2}}) = \pi(\tau_p)\pi(\theta_{\kappa,w})\pi(\theta_{\tau,w}) \tag{D.5}$$

where we have to assign some prior distributions to each hyperparameter in $\theta_{\mathbf{2}}$. The prior distributions used are INLA default priors which are Gaussian distributed with mean 0 and and variance $\sigma*^2$.

The LGMs SP_rb ad SP_rbp only differ by the number of hyperparameters in the latent vector. For SP_rb we have the vector $\theta_1 = [\mathbf{w}, \beta_0, \beta_1]$ for the unobserved process η_i now expressed as $\eta(\mathbf{s})$ where $\mathbf{s} = [s_1, ..., s_i]$, and for SP_rbp we have the vector $\theta_1 = [\mathbf{w}, \beta_0, \beta_1, \beta_2]$ for the unobserved process $\eta(\mathbf{s})$.

In the LGM for SP_rbpc we have introduced the spatially varying coefficient, $\tilde{\beta}_{j,i} = \beta_2 + \beta_{j,i}$, such as the observation models for SP_rbpc is defined as

$$\mathbf{y}*|\mathbf{w}, \mathbf{u}, \tau_c, \beta_0, \beta_1, \beta_2 \sim N(\eta, \tau_c^{-1}\mathbf{I}) \tag{D.6}$$

and the vector for the unobserved process η*$(\mathbf{s})$ becomes $\theta_1 = [\mathbf{w}, \mathbf{u}, \beta_0, \beta_1, \beta_2]$, where both **w** and **u** are GMRFs with prePIsion matrices as given in **eq. 3.25**. Within the Bayesian framework we assume a Gaussian prior to our intercept β_0, and we assume Gaussian priors for the coefficients of the two explanatory variables g and p, β_1 and β_3. This can be seen as

$$\mathbf{w}|\theta_{\tau,w}, \theta_{\kappa,w} \sim N(0, \mathbf{Q}^{-1}(\theta_{\tau,w}, \theta_{\kappa,w})) \tag{D.7}$$

$$\mathbf{u}|\theta_{\tau,u}, \theta_{\kappa,u} \sim N(0, \mathbf{Q}^{-1}(\theta_{\tau,u}, \theta_{\kappa,u})) \tag{D.8}$$

$$\beta_0 \sim N(\cdot, \cdot) \tag{D.9}$$
$$\beta_1 \sim N(\cdot, \cdot) \tag{D.10}$$
$$\beta_3 \sim N(\cdot, \cdot) \tag{D.11}$$

Appendix E

Results linear models

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-3608.0519	2352.6662	-1.53	0.1265
utm_east_z33	-0.0028	0.0005	-5.35	**0.0000**
utm_north_z33	0.0008	0.0004	2.17	0.0308
area_total	0.0083	0.0190	0.44	0.6610
gradient_1085	-9.0641	3.2009	-2.83	0.0050
gradient_basin	20.3458	3.8772	5.25	**0.0000**
gradient_river	12.6899	3.4224	3.71	**0.0003**
height_minimum	0.4048	0.2558	1.58	0.1149
height_hypso_10	0.6331	1.0686	0.59	0.5542
height_hypso_20	0.2194	2.4962	0.09	0.9300
height_hypso_30	-5.8039	4.0228	-1.44	0.1504
height_hypso_40	4.0449	4.7710	0.85	0.3974
height_hypso_50	-2.2236	5.0481	-0.44	0.6600
height_hypso_60	6.7968	5.7536	1.18	0.2387
height_hypso_70	-7.4592	5.5227	-1.35	0.1781
height_hypso_80	3.3739	3.7216	0.91	0.3656
height_hypso_90	-1.6005	1.4190	-1.13	0.2605
height_maximum	0.3079	0.2700	1.14	0.2553
length_km_basin	1.6873	4.1640	0.41	0.6857
length_km_river	-1.8657	2.5838	-0.72	0.4710
perc_agricul	-14.5546	8.7347	-1.67	0.0970
perc_bog	13.4708	8.2881	1.63	0.1054
perc_eff_lake	-30.4932	15.9756	-1.91	0.0575
perc_forest	-3.1948	4.9273	-0.65	0.5174
perc_glacier	23.8127	7.3291	3.25	**0.0013**
perc_lake	21.6914	10.9884	1.97	0.0495
perc_mountain	13.9847	5.3211	2.63	0.0091
perc_urban	24.6405	88.5146	0.28	0.7810

Table E.1: Results form LM1.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	3.6273	1.4401	2.52	0.0124
utm_east_z33	-0.0000	0.0000	-6.14	**0.0000**
utm_north_z33	0.0000	0.0000	2.32	0.0213
area_total	0.0000	0.0000	0.94	0.3502
gradient_1085	-0.0031	0.0020	-1.58	0.1155
gradient_basin	0.0121	0.0024	5.11	**0.0000**
gradient_river	0.0044	0.0021	2.10	0.0366
height_minimum	0.0001	0.0002	0.78	0.4340
height_hypso_10	-0.0000	0.0007	-0.02	0.9825
height_hypso_20	0.0003	0.0015	0.19	0.8482
height_hypso_30	-0.0021	0.0025	-0.85	0.3955
height_hypso_40	0.0010	0.0029	0.36	0.7201
height_hypso_50	-0.0008	0.0031	-0.27	0.7881
height_hypso_60	0.0017	0.0035	0.49	0.6261
height_hypso_70	-0.0019	0.0034	-0.57	0.5690
height_hypso_80	0.0013	0.0023	0.59	0.5564
height_hypso_90	-0.0009	0.0009	-1.00	0.3189
height_maximum	0.0002	0.0002	1.22	0.2249
length_km_basin	-0.0001	0.0025	-0.03	0.9741
length_km_river	-0.0015	0.0016	-0.94	0.3464
perc_agricul	-0.0103	0.0053	-1.93	0.0554
perc_bog	0.0131	0.0051	2.59	0.0102
perc_eff_lake	-0.0178	0.0098	-1.83	0.0692
perc_forest	-0.0003	0.0030	-0.09	0.9295
perc_glacier	0.0190	0.0045	4.24	**0.0000**
perc_lake	0.0227	0.0067	3.37	**0.0009**
perc_mountain	0.0126	0.0033	3.87	**0.0001**
perc_urban	0.0039	0.0542	0.07	0.9429

Table E.2: Results form LM1_c_log.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.3172	2291.9406	0.00	0.9975
utm_east_z33	-0.0010	0.0006	-1.71	0.0889
utm_north_z33	0.0001	0.0004	0.18	0.8559
area_total	0.0050	0.0178	0.28	0.7779
gradient_1085	-8.4243	3.0028	-2.81	0.0054
gradient_basin	19.2175	3.6400	5.28	**0.0000**
gradient_river	11.3706	3.2165	3.54	**0.0005**
height_minimum	0.3760	0.2398	1.57	0.1183
height_hypso_10	0.5181	1.0020	0.52	0.6056
height_hypso_20	0.5826	2.3410	0.25	0.8037
height_hypso_30	-4.2425	3.7809	-1.12	0.2630
height_hypso_40	2.3368	4.4825	0.52	0.6026
height_hypso_50	-1.2307	4.7356	-0.26	0.7952
height_hypso_60	4.2969	5.4111	0.79	0.4279
height_hypso_70	-4.5343	5.2020	-0.87	0.3843
height_hypso_80	2.4110	3.4929	0.69	0.4907
height_hypso_90	-1.5093	1.3304	-1.13	0.2577
height_maximum	0.3123	0.2532	1.23	0.2186
length_km_basin	2.3986	3.9056	0.61	0.5397
length_km_river	-2.1065	2.4227	-0.87	0.3855
perc_agricul	-14.0148	8.1892	-1.71	0.0883
perc_bog	13.8197	7.7702	1.78	0.0766
perc_eff_lake	-12.8636	15.2821	-0.84	0.4008
perc_forest	0.2595	4.6575	0.06	0.9556
perc_glacier	16.4143	6.9883	2.35	0.0197
perc_lake	10.4671	10.4816	1.00	0.3190
perc_mountain	11.0806	5.0135	2.21	0.0281
perc_urban	48.7612	83.0855	0.59	0.5578
avg_5	0.4215	0.0727	5.80	**0.0000**

Table E.3: Results form LM_cn.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	5.5281	1.4289	3.87	**0.0001**
utm_east_z33	-0.0000	0.0000	-2.81	0.0054
utm_north_z33	0.0000	0.0000	0.60	0.5499
area_total	0.0000	0.0000	0.82	0.4113
gradient_1085	-0.0028	0.0019	-1.47	0.1419
gradient_basin	0.0115	0.0023	5.08	**0.0000**
gradient_river	0.0037	0.0020	1.85	0.0655
height_minimum	0.0001	0.0001	0.72	0.4727
height_hypso_10	-0.0001	0.0006	-0.12	0.9048
height_hypso_20	0.0005	0.0015	0.33	0.7406
height_hypso_30	-0.0013	0.0024	-0.54	0.5890
height_hypso_40	0.0001	0.0028	0.05	0.9573
height_hypso_50	-0.0003	0.0030	-0.10	0.9167
height_hypso_60	0.0004	0.0034	0.12	0.9048
height_hypso_70	-0.0004	0.0032	-0.12	0.9044
height_hypso_80	0.0008	0.0022	0.38	0.7015
height_hypso_90	-0.0008	0.0008	-0.99	0.3241
height_maximum	0.0002	0.0002	1.29	0.1988
length_km_basin	0.0003	0.0024	0.12	0.9050
length_km_river	-0.0016	0.0015	-1.07	0.2849
perc_agricul	-0.0100	0.0051	-1.96	0.0511
perc_bog	0.0133	0.0048	2.75	**0.0064**
perc_eff_lake	-0.0086	0.0095	-0.90	0.3688
perc_forest	0.0015	0.0029	0.53	0.5942
perc_glacier	0.0151	0.0044	3.47	**0.0006**
perc_lake	0.0168	0.0065	2.56	0.0110
perc_mountain	0.0111	0.0031	3.55	**0.0005**
perc_urban	0.0166	0.0518	0.32	0.7494
avg_5	0.0002	0.0000	4.89	**0.0000**

Table E.4: Results form LM_cn_log.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2377.4479	2011.7625	-1.18	0.2385
utm_east_z33	-0.0008	0.0005	-1.61	0.1086
utm_north_z33	0.0004	0.0003	1.25	0.2120
area_total	0.0104	0.0155	0.67	0.5026
gradient_1085	-12.6459	2.6679	-4.74	**0.0000**
gradient_basin	10.9081	3.3109	3.29	**0.0011**
gradient_river	11.1778	2.8125	3.97	**0.0001**
height_minimum	0.2210	0.2114	1.05	0.2970
height_hypso_10	0.5798	0.9057	0.64	0.5227
height_hypso_20	0.6370	2.1368	0.30	0.7659
height_hypso_30	-3.2646	3.3387	-0.98	0.3292
height_hypso_40	2.1548	3.8978	0.55	0.5809
height_hypso_50	-3.5669	4.1377	-0.86	0.3895
height_hypso_60	6.4810	4.7369	1.37	0.1726
height_hypso_70	-3.9800	4.5407	-0.88	0.3816
height_hypso_80	1.5938	3.0412	0.52	0.6007
height_hypso_90	-1.0075	1.1583	-0.87	0.3853
height_maximum	0.0866	0.2216	0.39	0.6964
length_km_basin	0.8780	3.4020	0.26	0.7966
length_km_river	-1.1627	2.1098	-0.55	0.5821
perc_agricul	-10.2626	7.1582	-1.43	0.1530
perc_bog	3.0659	6.8759	0.45	0.6561
perc_eff_lake	1.6950	13.3884	0.13	0.8994
perc_forest	-2.2442	4.0662	-0.55	0.5815
perc_glacier	1.7610	6.3021	0.28	0.7802
perc_lake	-8.5611	9.3694	-0.91	0.3618
perc_mountain	3.5402	4.4421	0.80	0.4263
perc_urban	6.5987	72.4328	0.09	0.9275
avg_5	0.0777	0.0746	1.04	0.2990
median_prec	0.6218	0.0714	8.71	**0.0000**

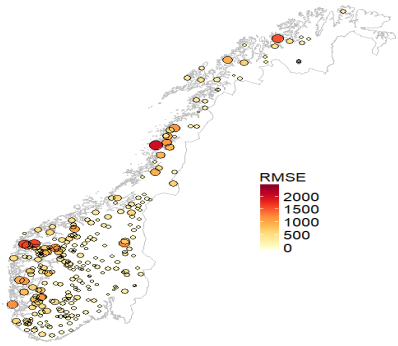
Table E.5: Results form LM_cnp.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.9990	1.2810	-0.78	0.4363
utm_east_z33	-0.0000	0.0000	-2.06	0.0405
utm_north_z33	0.0000	0.0000	1.73	0.0851
area_total	0.0000	0.0000	0.82	0.4105
gradient_1085	-0.0052	0.0015	-3.40	**0.0008**
gradient_basin	0.0040	0.0019	2.05	0.0416
gradient_river	0.0038	0.0016	2.37	0.0186
height_minimum	0.0000	0.0001	0.18	0.8539
height_hypso_10	0.0002	0.0005	0.35	0.7252
height_hypso_20	0.0001	0.0012	0.07	0.9449
height_hypso_30	-0.0006	0.0019	-0.31	0.7559
height_hypso_40	0.0002	0.0022	0.07	0.9419
height_hypso_50	-0.0016	0.0024	-0.67	0.5039
height_hypso_60	0.0025	0.0027	0.92	0.3600
height_hypso_70	-0.0008	0.0026	-0.32	0.7523
height_hypso_80	0.0003	0.0018	0.19	0.8520
height_hypso_90	-0.0004	0.0007	-0.67	0.5018
height_maximum	0.0001	0.0001	0.40	0.6884
length_km_basin	0.0006	0.0020	0.33	0.7423
length_km_river	-0.0013	0.0012	-1.05	0.2932
perc_agricul	-0.0086	0.0041	-2.09	0.0373
perc_bog	0.0027	0.0040	0.66	0.5076
perc_eff_lake	0.0020	0.0077	0.26	0.7976
perc_forest	-0.0015	0.0023	-0.62	0.5355
perc_glacier	0.0015	0.0037	0.42	0.6783
perc_lake	-0.0025	0.0055	-0.46	0.6485
perc_mountain	0.0038	0.0026	1.46	0.1444
perc_urban	-0.0254	0.0418	-0.61	0.5434
avg_5	0.0000	0.0000	0.06	0.9498
median_prec	0.8185	0.0714	11.46	**0.0000**

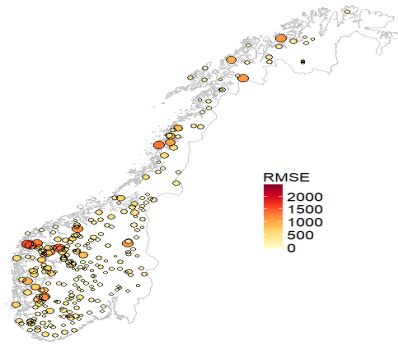
Table E.6: Results form LM_cnp_log.

Appendix F

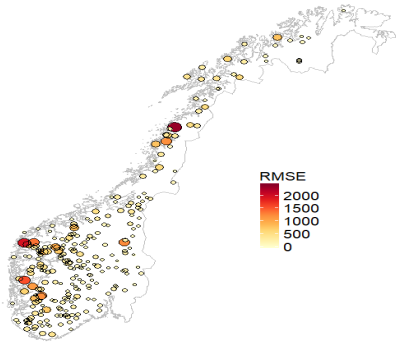
Additional results spatial models



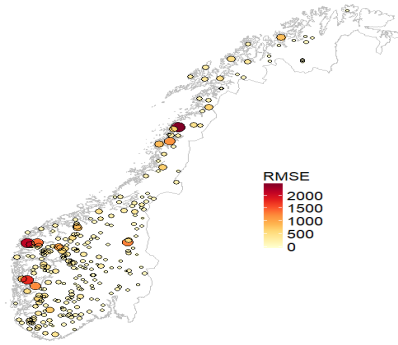
(a) SP_r RMSE score.



(b) SP_rb RMSE score.

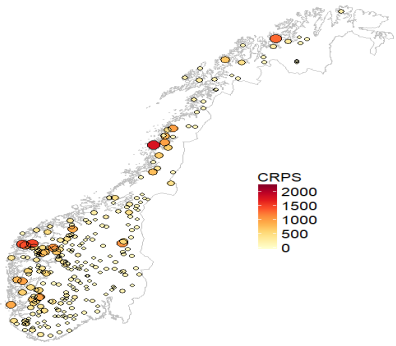


(c) SP_rbp RMSE score.

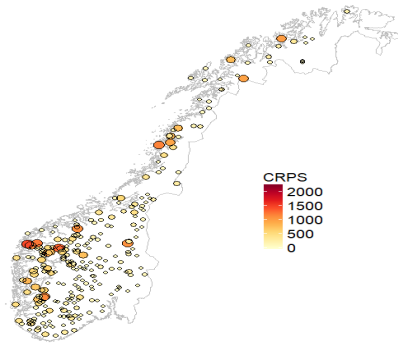


(d) SP_rbpc RMSE score.

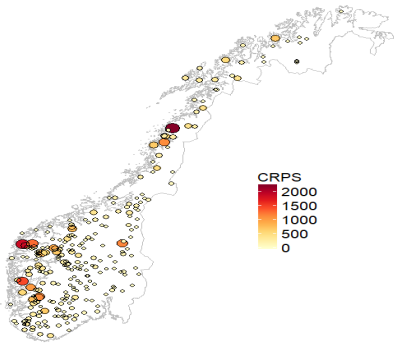
Figure F.1: Map showing RMSE scores from our spatial models.



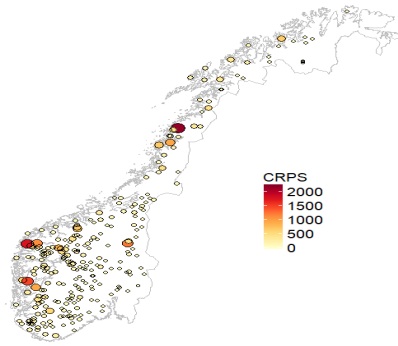
(a) SP_r CRPS score.



(b) SP_rb CRPS score.

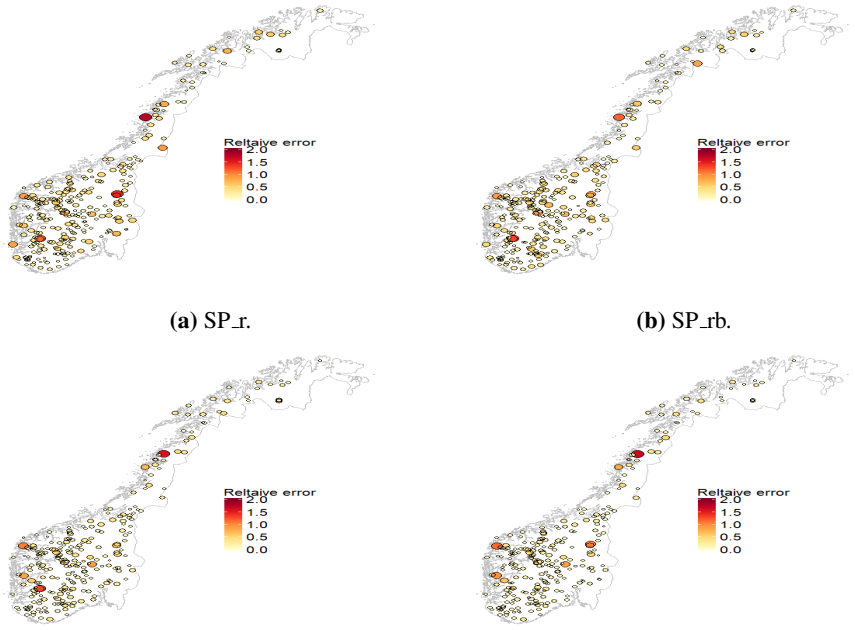


(c) SP_rbp CRPS score.



(d) SP_rbpc CRPS score.

Figure F.2: Map showing CRPS scores from our spatial models.



(a) SP_r.

(b) SP_rb.

(c) SP_rbp.

(d) SP_rbpc.

Figure F.3: Absolute relative error from our spatial models plotted in a map with point locations of the 266 catchments.

Appendix G

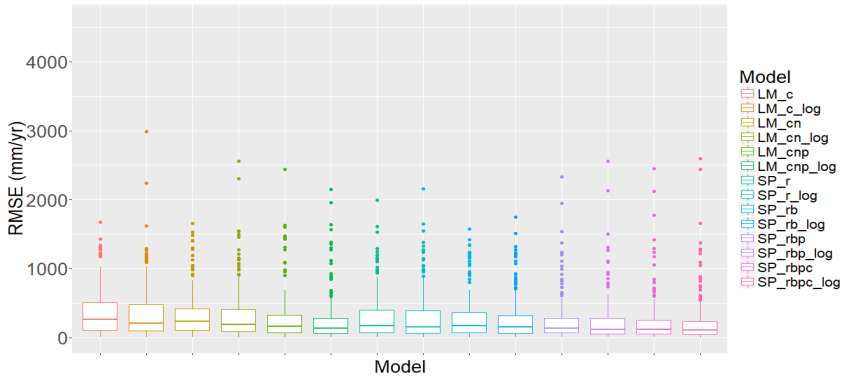
Transformed models

	Model	RMSE	CRPS
1	LM_c_log	332.24	1270.00
2	LM_cn_log	309.24	1281.00
3	LM_cnp_log	238.80	1315.00
4	SP_r_log	287.70	1309.10
5	SP_rb_log	259.90	1140.30
6	SP_rbp_log	221.10	1172.00
7	SP_rbpc_log	209.20	1143.00

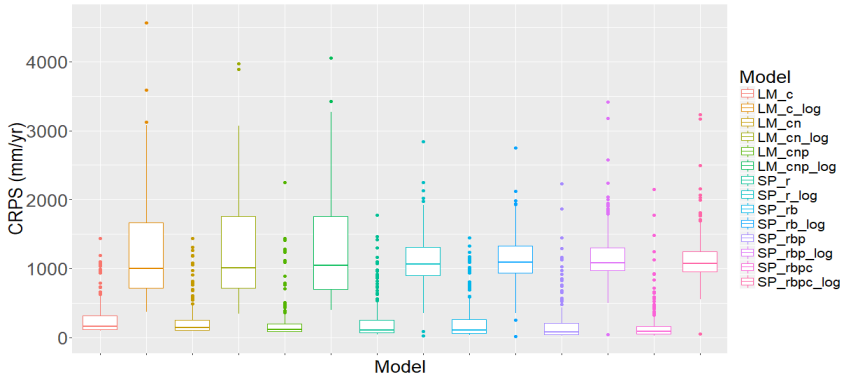
Table G.1: Evaluation of model performance for the log-transformed models. The model performance is evaluated with mean RMSE and mean CRPS score for the LOOCV predictions.

The models presented in chapter 4.1 and 4.3 where also fitted to the logarithm of runoff. This was done to explore if it could remove the observed heterogeneity in our residuals (see **fig. 5.9a** and **fig. 6.4**). Evaluation of predictive performance was done for the original scale, e.g. predictions of runoff were back-transformed.

A log transformation of the dependent variable median annual runoff for our models (**tab. 4.1** and **4.4**) was done as an attempt to remove heterogeneity in our data. The CRPS scores of the log-transformed models in **tab. G.1**, tells us that something is going terribly wrong as the CRPS value shows much higher values than the non-transformed models. The CRPS is calculated based on the whole posterior predictive distribution, and thus reviles more information about the posterior predictive distribution than RMSE which only evaluates the difference between predicted and observed value. If we have a look at **tab. G.1a** all RMSE values for all our models are plotted. It shows that there is not much difference in the transformed and non-transformed models for the RMSE value. If we look at **tab. G.1b** we see how there is a much larger difference in CRPS values within the linear and spatial models. The log-transformed linear models are worst in terms of CRPS score. The log-transformed spatial models do not perform much better, having much larger upper tails than the corresponding non-transformed models.



(a) RMSE values of LOOCV prediction for all our models.



(b) CRPS values of LOOCV prediction for all our models.

Figure G.1: Evaluation of model performance of all our models displayed in boxplots. **(a)** Shows the RMSE score and **(b)** shows the CRPS score.

Further we look at the residuals plotted against the predicted median annual runoff in **fig. G.2**. The log-transformed linear models (**fig. G.2a**) shows a cone shape that indicate heterogeneity and we also observe some observations that are much larger than the residual value we saw for the non-transformed linear models in **fig. 5.9a**. The log-transformed spatial models in **fig. G.2b** shows a cone shape as we did for the log-transformed linear models, where we also have some very large residuals that are much larger than the outliers seen in **fig. 6.4** for the non-transformed models.

To illustrate what happens when log-transforming the models we have plotted the prediction interval (PI) of all our models. By looking at the PIs for our linear models with catchment characteristics as the covariates (LM_c and LM_c_log) in **fig. G.6** we see how the PIs for the non-transformed model (**fig. G.6a**) are quite similar for all observations, and also that the range of the PIs are around 2000 mm/yr. For the transformed model (**fig. G.6b**) the PIs differ a lot, for small observations the PIs are more narrow with a range of approximately 800 mm/yr, but for larger observations the PIs are much larger, and the

(a) Residuals of LM_log models.

(b) Residuals of SP_log models.

Figure G.2: Residuals plotted against LOOCV predicted values for the log-transformed models. **(a)** is for the log-transformed linear model, **(b)** is the log-transformed spatial models.

largest PIs have a range of almost 5000 mm/yr. These large PIs get punished in the CRPS and they give us a large CRPS value. The same can be said about the linear models with catchment characteristics and a spatial dependency term (LM_cn and LM_cn_log) plotted in **fig. G.7**. For the log-transformed model (**fig. G.7b**) the PIs are small for small observations of runoff and for the large observations is has a range of approximately 3000 mm/yr. When precipitation is included, in addition to catchment characteristics and a spatial term in the linear models (LM_cnp and LM_cnp_log), the overall PIs get a smaller range, as seen in **fig. G.8**, but the log-transformed model remains to have to wide PIs for large observations.

The PIs for the spatial models are plotted in **fig. G.9**, **G.10**, **G.11** and **G.12**. Here we see that the PIs for our log-transformed models are much too wide, as it ranges from -4000 and up to 4000 it is not a quality we want. We also notice how the PIs of the non-transformed models are much more narrow than what any other model in this thesis show, and although some observations are not covered by its PI, most observations are.

To further look at what happens with the log-transformed models we have chosen to look at two catchments posterior predictive distribution. The two catchments are namely *Fiskum* (field ID 515) and *Risevatn* (field ID 1440). Both are plotted in a map in **fig. G.3**. They were chosen based on the residual value from the log-transformed models. *Fiskum* has a residual value of -250 mm/yr, while **fig. G.3** have a residual value of -1612 mm/yr.

We first look at the posterior predictive distribution of our linear models in **fig. G.4**. The distribution for catchment named *Fiskum* (field ID 515) plotted in **fig. G.4a** and **G.4c** shows a much wider standard deviation for the log-transformed than the non-transformed. Also the mean of the non-transformed are much smaller than the mean of the transformed model. For *Risevatn* (field ID 1440) both the non-transformed (**fig. G.4b**) and the transformed (**G.4c**) has a mean much larger than observed. For *Fiskum* we observe that the posterior predictive distribution is much wider for the transformed model, than for the non-transformed model. The posterior predictive distribution is also much larger for *Risevatn* than for *Fiskum*.

The posterior predictive distribution of our spatial models is seen in (**fig. G.5**). For catchment number 515 the non-transformed models (**fig. G.5a**) perform very well in terms

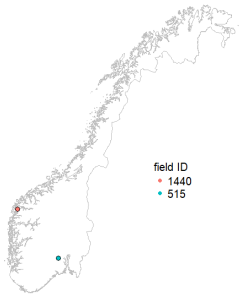
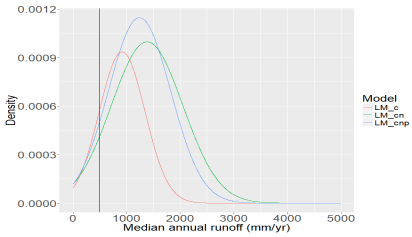
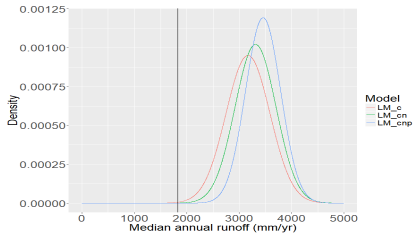


Figure G.3: Map showing the point locations of the two catchments used to illustrate posterior predictive distribution of our models.

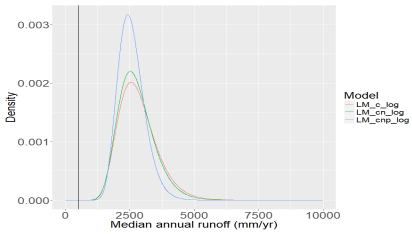
of prediction versus observed, there is a difference between the different models within this catchment. We see that the SP_rbpc (with spatially varying covariate) has the smallest standard deviation and that SP_r (with spatially random field) has a larger standard deviation of the spatial models. The log-transformed models (**fig. G.5c**) shows that SP_rbp and SP_rbpc (both with precipitation) are similar, and so are the two models SP_r and SP_rb (both without precipitation) where the standard deviation is much larger. For catchment with field ID 1440 (**fig. G.5b** and **G.5d**) the models do not fit well with the observed median annual runoff. Here the SP_rbpc has a much larger standard deviation than the other models. The mean and standard deviation for the models at this catchment also seem to be quite different, and the standard deviation of the transformed models are much larger than the standard deviation of the non-transformed models.



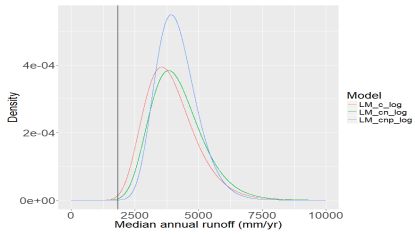
(a) Field ID 515.



(b) Field ID 1440.

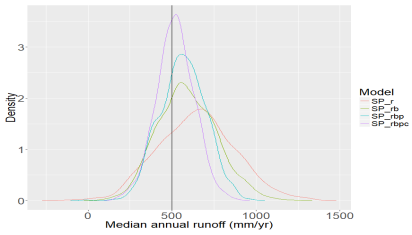


(c) Field ID 515.

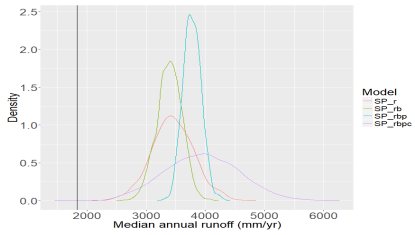


(d) Field ID 1440.

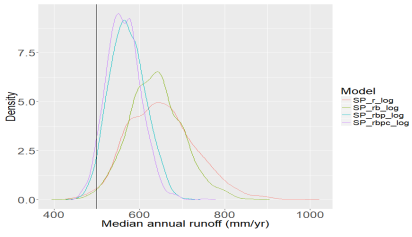
Figure G.4: Posterior predictive distribution for *Fiskum* (field ID 515) and *Risevatn* (field ID 1440) with the linear models. The black line represents the observed value of runoff.



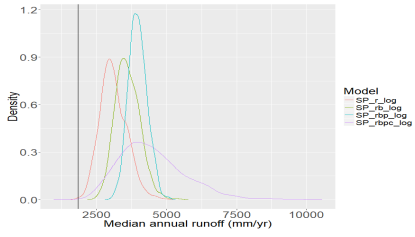
(a) Field ID 515.



(b) Field ID 1441.

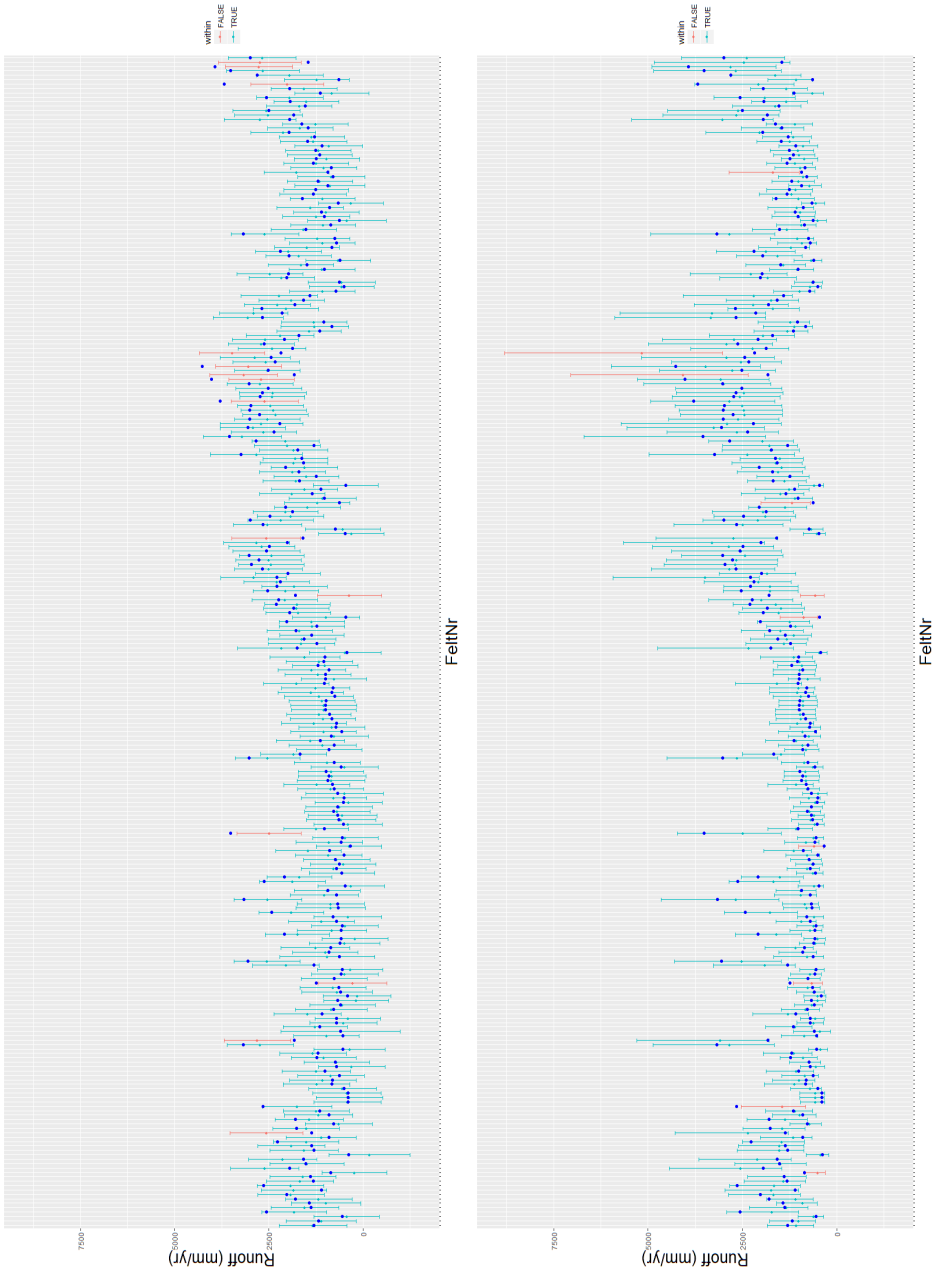


(c) Field ID 515.



(d) Field ID 1440.

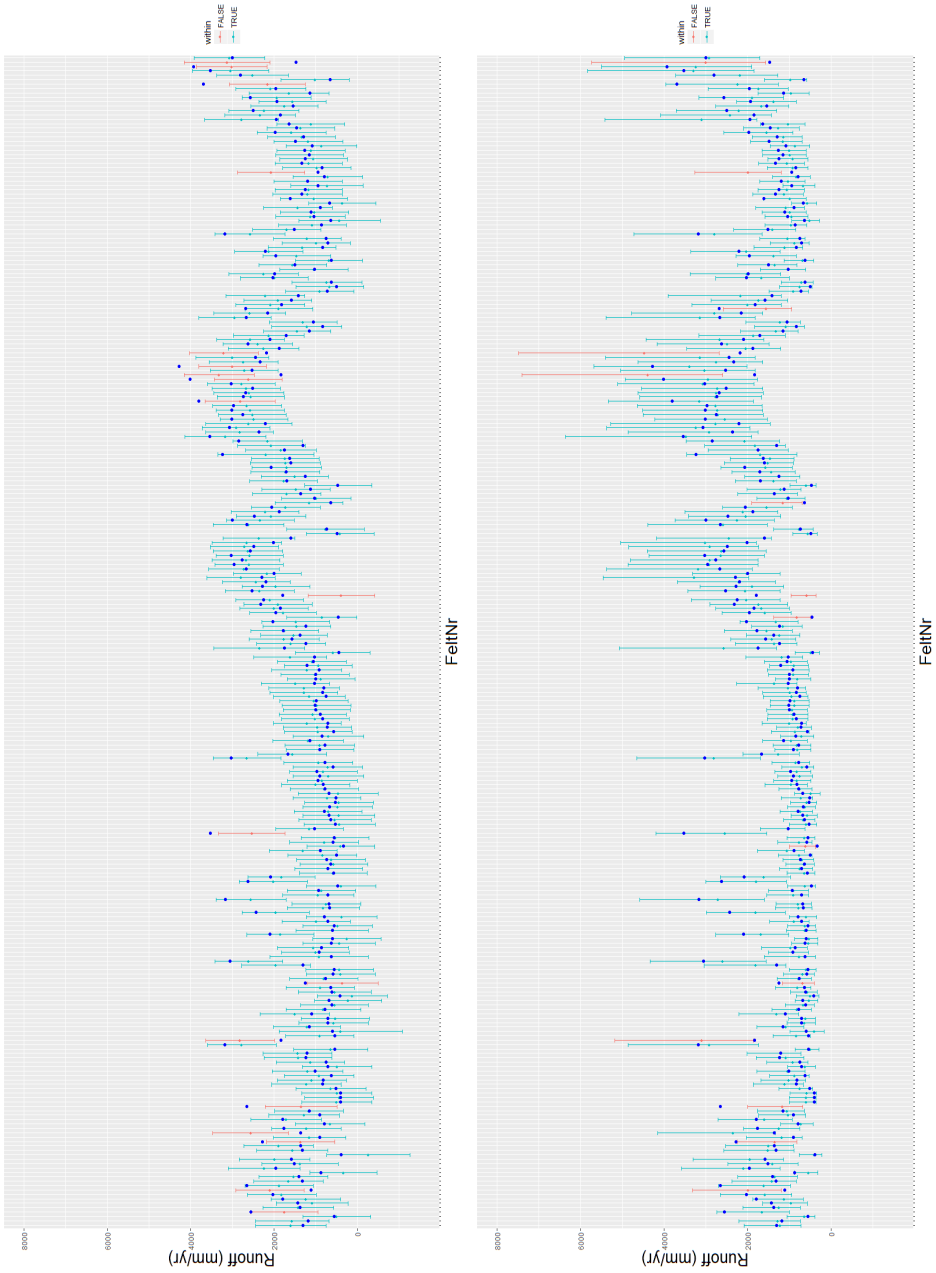
Figure G.5: Posterior predictive distribution for *Fiskum* (field ID 515) and *Risevatn* (field ID 1440) with the spatial models. The black line represents the observed value of runoff.



(a) LM_c 95% PI.

(b) LM_c_log 95% PI.

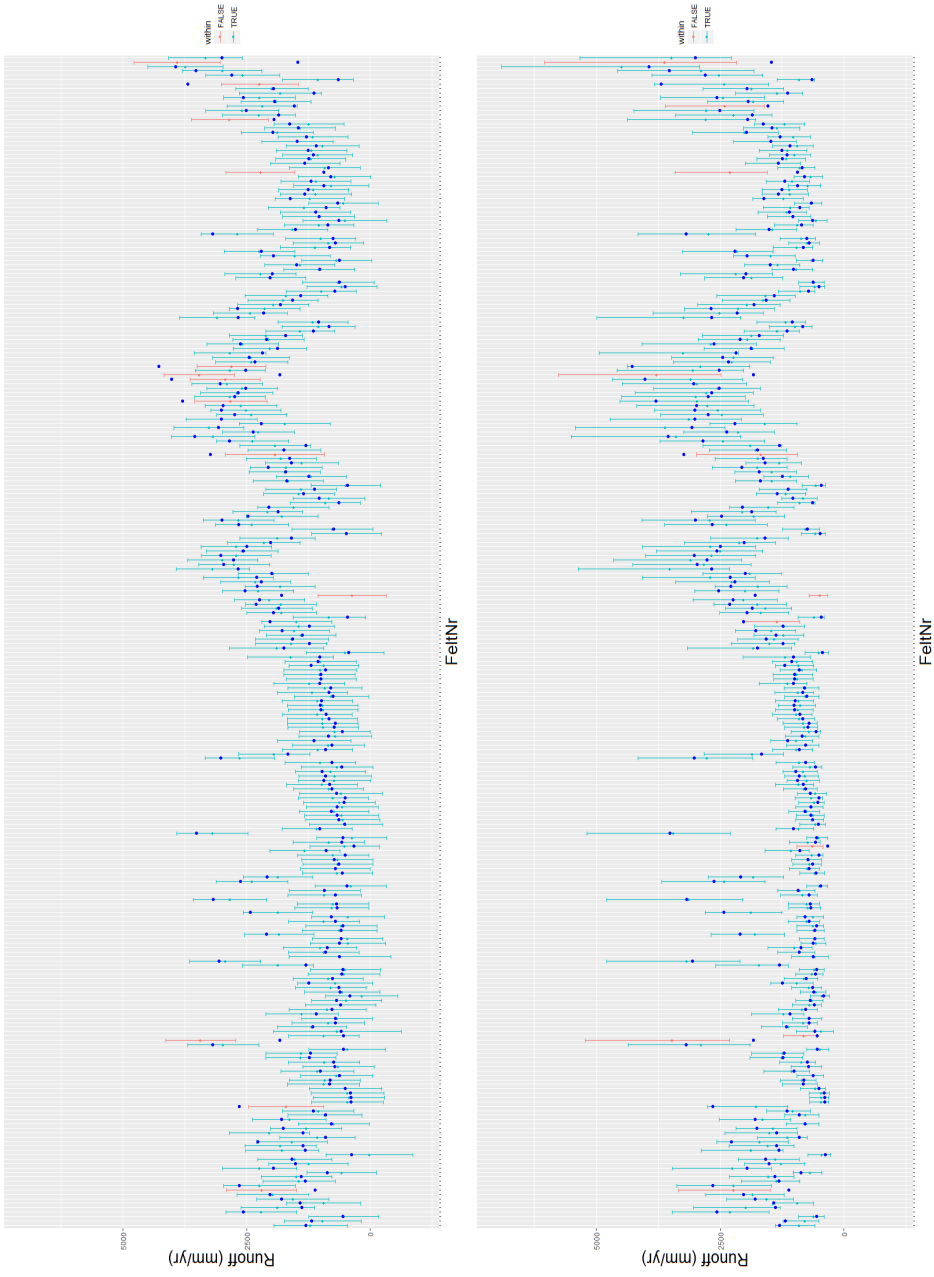
Figure G.6: Plots of the 95% prediction interval for the models LM_c and LM_c_log.



(a) LM_cn 95% PI.

(b) LM_cn_log 95% PI.

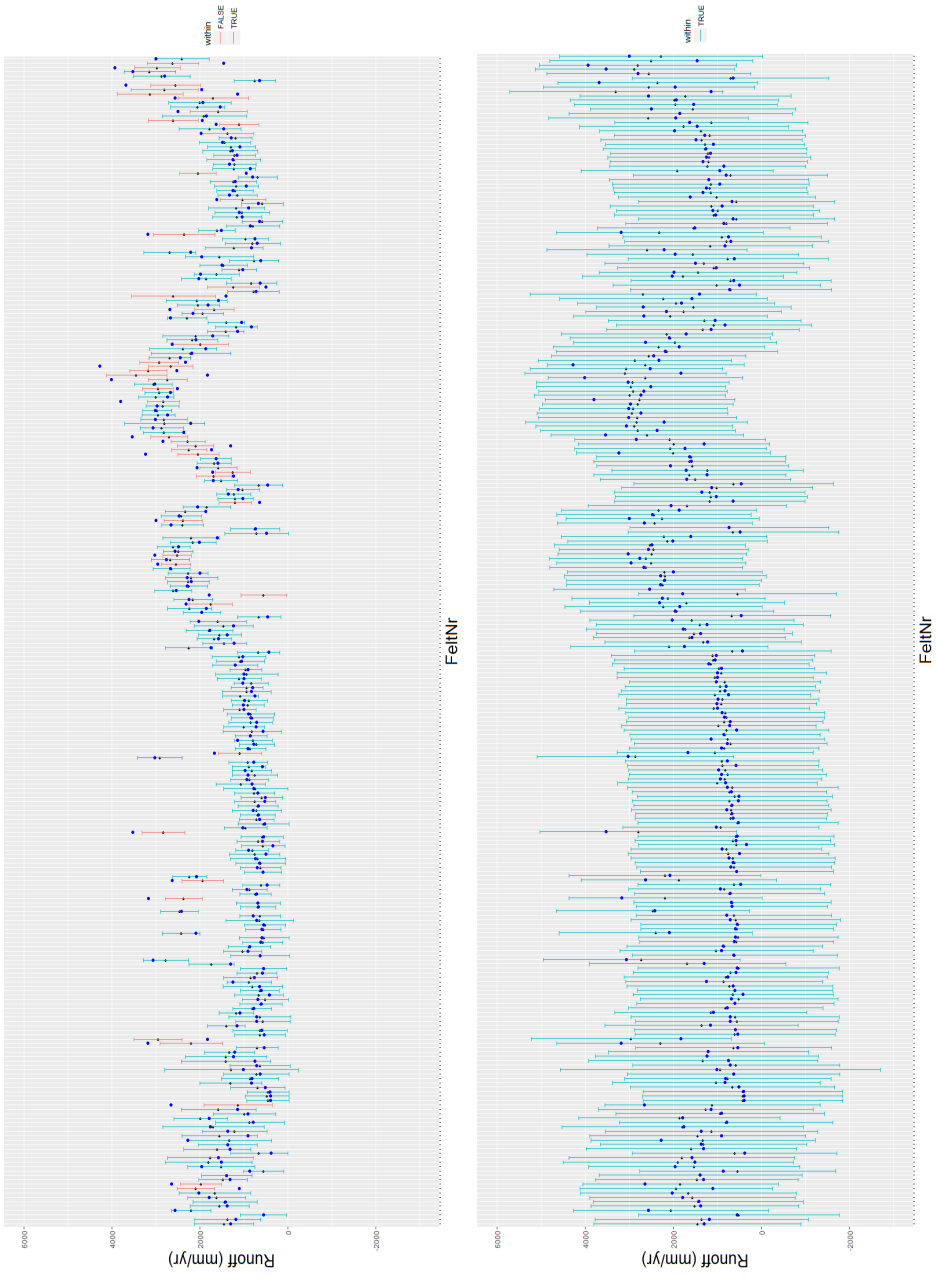
Figure G.7: Plots of the 95% prediction interval for the models LM_cn and LM_cn_log.



(a) LM_cnp 95% PI.

(b) LM_cnp_log 95% PI.

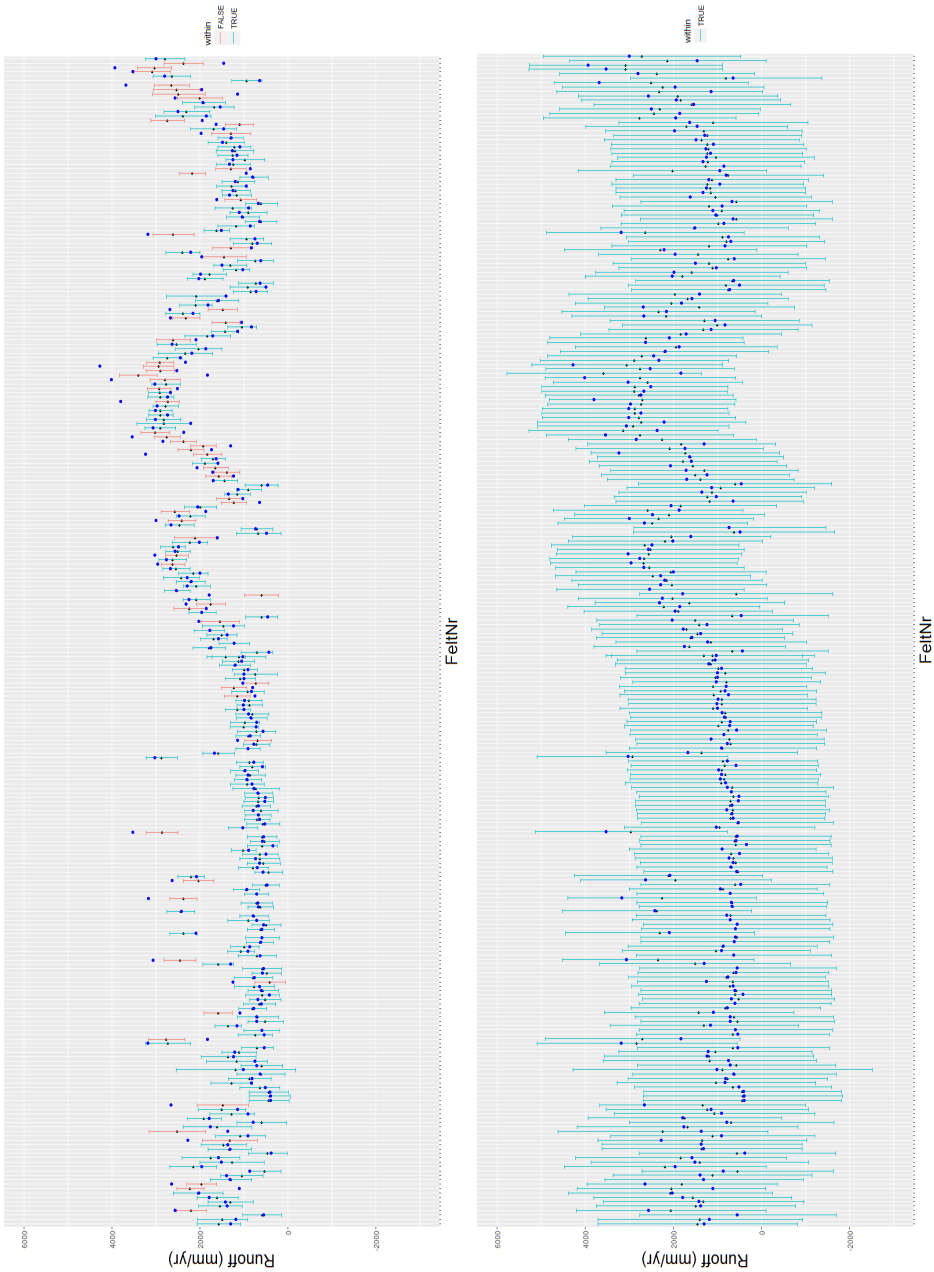
Figure G.8: Plots of the 95% prediction interval for the models LM_cnp and LM_cnp_log.



(a) SP_r 95% PI.

(b) SP_r_log 95% PI.

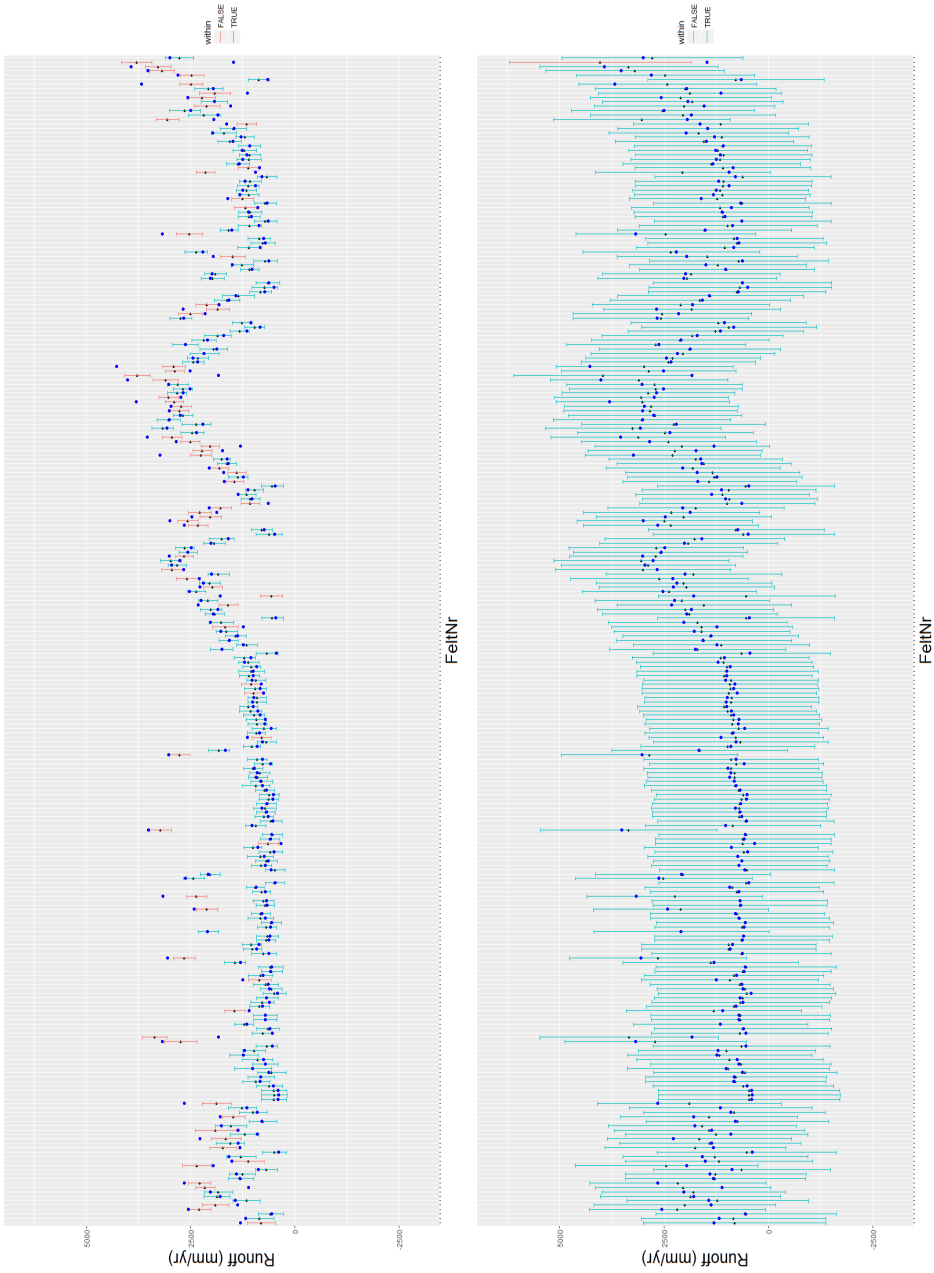
Figure G.9: Plots of the 95% prediction interval for the models SP_r and SP_r_log.



(a) SP_rb 95% PI.

(b) SP_rb_log 95% PI.

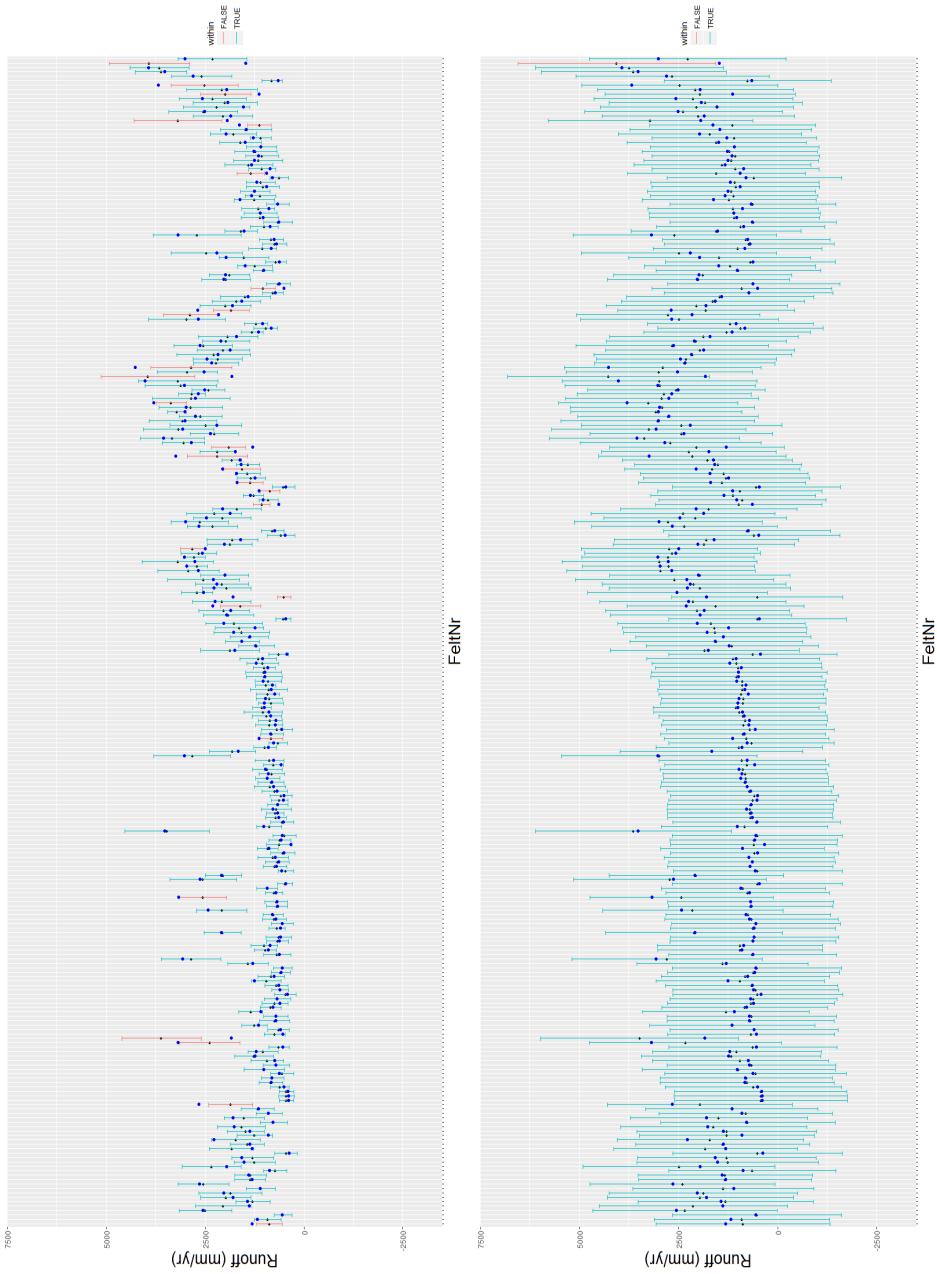
Figure G.10: Plots of the 95% prediction interval for the models SP_rb and SP_rb_log.



(a) SP_rbp 95% PI.

(b) SP_rbp_log 95% PI.

Figure G.11: Plots of the 95% prediction interval for the models SP_rbp and SP_rbp_log.



(a) SP_rbpc 95% PI.

(b) SP_rbpc_log 95% PI.

Figure G.12: Plots of the 95% prediction interval for the models SP_rbpc and SP_rbpc_log.