Doctoral theses at NTNU, 2020:268

Thea Julie Thømt Roksvåg

Bayesian geostatistical two field models for combining data sources and exploiting short records

Applied to annual runoff interpolation in Norway

ISBN 978-82-326-4884-9 (printed ver.)
ISBN 978-82-326-4885-6 (electronic ver.)

Doctoral theses at NTNU, 2020:268

NTNU
Norwegian University of
Science and Technology
Thesis for the degree of
Philosophiae Doctor
Faculty of Information Technology
and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

NTNU

NTNU
Norwegian University of
Science and Technology

Thea Julie Thømt Roksvåg

Bayesian geostatistical two field models for combining data sources and exploiting short records

Applied to annual runoff interpolation in Norway

Thesis for the degree of Philosophiae Doctor

Trondheim, September 2020

Norwegian University of Science and Technology
Faculty of Information Technology
and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology
and Electrical Engineering
Department of Mathematical Sciences

ISBN 978-82-326-4884-9 (printed ver.)
ISBN 978-82-326-4885-6 (electronic ver.)
ISSN 1503-8181

Doctoral theses at NTNU, 2020:268

Printed by Skipnes Kommunikasjon AS

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of Philiosophiae Doctor (PhD) at the Norwegian University of Science and Technology (NTNU). The work was carried out at the Department of Mathematical Sciences during the years 2016-2020 and was funded by the Research Council of Norway.

First of all I would like to thank my supervisor Ingelin Steinsland for excellent guidance through five years, first through my master's thesis work and then through this PhD. She also encouraged me to apply for this PhD position, and for that I am very grateful. Next, I would like to thank my co-supervisor Kolbjørn Engeland for sharing his knowledge about hydrology, for valuable discussions and for interesting visits at NVE in Oslo. I would also like to thank Stein Beldring for giving me the opportunity to work more closely with NVE projects.

The working environment at IMF has made the years as a PhD student enjoyable, and for that I would like to thank my coworkers and friends at NTNU. A particular thanks to my friend and colleague Maria who has been my office-mate almost the whole period. We have had many memorable conference trips and shared ups and downs throughout our PhDs. I would also like to mention my friends, Silje and Sigrid who really have been there for me these four years. Finally, I would like to thank Torstein, my parents Margrete and Andor, and my sister Anna for always being supportive.

Thea Roksvåg

Trondheim, June 2020

Thesis outline

Introduction

Paper A **A geostatistical two field model that combines point observations and nested areal observations, and quantifies long-term spatial variability - A case study of annual runoff predictions in the Voss area**

Thea Roksvåg, Ingelin Steinsland and Kolbjørn Engeland

In revision in a statistics journal

Paper B **Estimation of annual runoff by exploiting long-term spatial patterns and short records within a geostatistical framework**

Thea Roksvåg, Ingelin Steinsland and Kolbjørn Engeland

Accepted for publication in Hydrology and Earth System Sciences

Paper C **A geostatistical spatially varying coefficient model for mean annual runoff that incorporates process-based simulations and short records**

Thea Roksvåg, Ingelin Steinsland and Kolbjørn Engeland

Technical note **An overview of a geostatistical two field model for interpolation of environmental variables**

Thea Roksvåg

Introduction

Introduction

Spatial statistics is the field of statistics analyzing data that are connected to geographical locations (Cressie, 1993; Gelfand et al., 2010; Banerjee et al., 2004). It is used within a range of scientific fields, e.g. for modeling the spatial distribution of diseases (Moraga, 2019; Berke, 2004; Clements et al., 2006), mapping weather phenomena and natural resources (Gandin, 1960; Kyriakidis et al., 2001; Skøien et al., 2006; Goovaerts, 1997; Paciorek and Schervish, 2006), mapping the behavior or habitat of animal species (Sicacha Parada et al., 2020; Bellier et al., 2010; Jullum et al., 2020) or for mapping the spatial distribution of demographic variables like income level, birth rate and child mortality across geographical regions (Wakefield et al., 2017; Kerry et al., 2010).

A common problem in spatial statistics is to predict a target variable at locations where there are no observations. To make predictions, relevant observations are gathered from nearby locations. With the development of GPS, satellite technology, Internet and citizen science, the availability of information is larger than ever. This means that it often is possible to combine several data sources when building a spatial model. However, this comes with challenges. The spatial observations might come in different forms, where some are point referenced, some are areal referenced and some are connected to grid units. Some data are aggregated and/or include information about averages over regions. There can also be large differences in how the data are collected and the measurement uncertainty associated with them: The uncertainty of a satellite observation is probably very different from the uncertainty of an observation reported by a person contributing to citizen science. Using different data types together in a spatial model, requires that the data are connected in a mathematically consistent way that gives an appropriate uncertainty quantification for the predicted quantity.

The variable of interest might also have other important properties that the researcher want to include the spatial model. For example can the target variable be subject to constraints, e.g. it follows preservation laws for mass and energy. Another possibility is that the variable follows repeated trends

over time, due to for example topography or other constant, underlying factors. Including such dependency structures and constraints in a spatial model is often computationally expensive.

In this thesis, we build Bayesian spatial models. The goal of Bayesian inference is to assess the posterior distribution of the target variable, given a set of observations (Gelman et al., 2004; Gamerman and Lopes, 2006). The benefits of using the Bayesian framework is that it enables a full uncertainty quantification for both the target variable and its underlying model parameters. It also makes it possible to include expert knowledge about e.g. observation uncertainties through prior distributions. However, the drawback is that Bayesian models often provide slow inference and predictions. A main goal of the thesis is to develop spatial models that account for different data types and data structures, but that also are computationally feasible for operational use.

The spatial models we present in our work, are motivated by challenges from hydrology and runoff estimation. Runoff is a hydrological variable that is driven by precipitation: When it rains or snows, some of the precipitation evaporates, some of it is stored (temporarily) as snow while the rest flows towards a river as runoff, either on the ground surface as *surface runoff* or within the soil as *subsurface runoff or interflow*. This process can be expressed through a basic water balance equation given by

$$P = Q + E + S, \tag{1}$$

where P is precipitation, Q is runoff, E is evapotranspiration and S is the change in stored water (determined by processes like snow melt and snow accumulation) (WMO, 1992). In our work, we mostly consider annual runoff. Annual runoff is typically given for *hydrological* years that are defined such that we can neglect the storage component S in the water balance equation. In this definition it is assumed that the snow that came during the hydrological year has melted before the beginning of the next hydrological year. With $S \approx 0$, we see that runoff can be observed through precipitation and evaporation data, two data types that are point referenced. However, it is more common to observe runoff through areal referenced streamflow measurements. These contain information about the average runoff within a *catchment* area. In addition is annual runoff a mass

conserved variable restricted by water balance constraints. Hence, runoff is a spatial variable that has many of the properties discussed above.

In our studies, we consider Norwegian annual runoff data. In Norway more than 90 % of the electrical energy produced comes from hydroelectricity (NVE, 2020), and runoff estimates and forecasts are needed to plan the hydropower production. Estimates of *mean annual runoff* give information about the long-term water availability across the country. This can be useful when determining the best location for building hydropower plants. Furthermore, it is important to estimate runoff and runoff extremes for flood security reasons: Infrastructure should be built such that it can handle floods with return periods of e.g. 100 or 200 years. Existing models used by the Norwegian water resources and energy directorate (NVE) have shown that mean annual runoff is one of the most essential predictors for runoff extremes (Engeland et al., 2020). In addition are (annual) runoff estimates used in the planning of domestic, agricultural and industrial water supply all over the world. The variability in annual runoff is also a key for understanding runoff's sensitivity to driving climatic factors in today's climate and can be used to make inference about the runoff variability for future climates (Blöschl et al., 2013).

Although runoff is of interest for a variety of reasons, most catchments in the world lack runoff measurements, and runoff must be estimated. This defines a key challenge in hydrology known as the prediction in ungauged basins problem (PUB) (Blöschl et al., 2013). This thesis contains three papers (Paper A-C) where all of them address the PUB problem. The papers are closely linked, but each contains a new contribution towards finding a better strategy for (annual) runoff interpolation.

In our work with runoff estimation, it was one problem that stood out as particularly interesting: The problem of exploiting short records of runoff (Vogel and Stedinger, 1985; Blöschl et al., 2013). For hydrologists it is often of interest to estimate the long-term behavior of runoff, e.g. the mean annual runoff. The long-term behavior, typically the behavior over a 30 year period, says something about the climate in the study area and the overall water availability. However, in some catchments there only exist 1, 3, 10 or 28 years of data when it is of interest to have 30 annual observations. These data records are what hydrologists refer to as short

records.

In Norway, the Norwegian Water Resources and Energy Directorate (NVE) are responsible for making mean annual runoff maps. In the latest versions of this map, short records have been omitted from the analysis (Beldring et al., 2002). The reason is that the short records have been considered as too uncertain and/or too complicated to include in an analysis with the existing methods for runoff interpolation. However, omitting the short records represents a reduction in the number of observation locations of more than 50 % in Norway. It is also known that the annual runoff in Norway, and in many other countries, tends to follow repeated spatial patterns over time suggesting that there can be a lot information in a tiny piece of data. Hence, the there is a potential of improving the existing geostatistical models in terms of exploiting short data records.

Motivated by the challenge of exploiting short records, we developed a Bayesian spatial model for annual runoff that models several years of runoff simultaneously with two spatial fields: One that represents the long-term average spatial variability of runoff, i.e. the climate, and one that represents year-specific spatial variability. The spatial model borrows strength in time through the climatic field, and can be though of as something between a spatial model and a spatio-temporal model.

During our work we have seen that the proposed two field model has several interesting properties that we present throughout the thesis. The study of the two field model can hence be considered a main statistical contribution of our work. To illustrate that the two field model can be used outside the field of annual runoff interpolation, the thesis includes a Technical note that focuses on more general properties of the model. Here, we also present examples of other environmental variables for which the model can be useful.

In addition to the study of the two field model, the thesis explores methods for combining different data sources in a spatial model. We consider how observations of different spatial support (point and areal observations of runoff) can be combined in a mathematically consistent way that accounts for basic preservation laws. We also explore how simulations from a physics-based hydrological model can be incorporated into a geostatistical

framework.

A main motivation behind the work was to develop models that were computationally feasible for operational use. This is not trivial when suggesting Bayesian models that include two spatial fields, models with areal referenced observations and models that consider up to 30 years of runoff simultaneously. To solve the computational problems, we use two relatively recent developments: (i) Integrated nested Laplace approximations (INLA) and (ii) the stochastic partial differential equation (SPDE) approach to spatial modeling (Rue et al., 2009; Lindgren et al., 2011). While INLA is a tool for making approximate inference on a specific class of Bayesian models, the SPDE approach is used to approximate Gaussian random fields (GRFs) by Gaussian markov random fields (GRMFs) through solving a SPDE numerically. These approximations come with computational benefits that are crucial to fit our suggested models.

While not a main topic of the thesis, we also wanted the models to be easy to handle in terms of prior specification. The idea is that this will contribute to making the models more available for operational use for e.g. hydrologists and other users that are not experts in statistics. Based on this we have used interpretable priors for most of the parameters, more specifically the recently developed PC priors suggested by Simpson et al. (2017). PC priors make it possible for the user to express prior believes through quantiles and probabilities in a simple manner. Knowledge-based priors are also used in this thesis to weight observation types differently in our spatial models, based on what we know about measurement uncertainties across data types.

Before describing our contributions, we give an overview of spatial statistics, GRFs, GMRFs, INLA, the SPDE approach, PC priors and other relevant statistical models and tools that are used to build our hydrological models. This can be found in Section 1. Next, in Section 2 we present the challenges that exist in runoff modeling that have motivated our work. Here, we also refer to how other researchers have addressed these issues and describe our contributions towards the field. Finally, in Section 3 we summarize the work, paper by paper, and suggest how the models can be used, also outside the field of runoff interpolation, and for further research.

Notation

In this introduction we use the following notation unless otherwise specified: Bold type lower case letters refer to vectors, while bold higher case letters refer to matrices. For example, $\boldsymbol{y}$ denotes a set of observations $(y_1, ..., y_n)$ while $\boldsymbol{\Sigma}$ refers to a covariance matrix with matrix element (i, j) given by $\Sigma_{i,j}$. To denote a spatial location in $\mathcal{R}^2$, we use $\boldsymbol{u}$ which is a vector containing an x-coordinate and a y-coordinate. For probability density functions, we use $\pi(\cdot)$, and to express conditional probabilities we use $\pi(\cdot|\cdot)$. An observation likelihood given conditional on some variables $\boldsymbol{x}$ and parameters $\boldsymbol{\theta}$ is for example written as $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$.

1 Methodological background

1.1 Spatial statistics

In this thesis we mostly consider models for spatial data that are continuously indexed, often referred to as geostatistical models. Geostatistics has it origins at L'École des Mines in France, where Georges Mathéron developed models for predicting the outcome of mining operations (Mathéron, 1955, 1963). His models were inspired by the work of D.G. Krige who used statistical methods for mineral exploration (Krige, 1966). Independently of Mathéron, the Swedish statistician Bertil Matérn established geostatistical models for forestry (Matérn, 1960). Here, we find the development of the Matérn correlation function that is commonly used to model correlation between locations.

A central assumption in spatial statistics is that locations that are close in space have more in common than locations that are far apart. This follows Tobler's first law of geography that states that *"everything is related to everything else, but near things are more related than distant things"* (Tobler, 1970). Assuming $\{x(\boldsymbol{u}); \boldsymbol{u} \in \mathcal{R}^2\}$ is a spatial process of interest, Tobler's law can be included in a spatial model by modeling how the dependency between $x(\boldsymbol{u}_i)$ and $x(\boldsymbol{u}_j)$ decays with distance for any two locations $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$. Such dependencies can be modeled by Gaussian random fields

(GRFs), a commonly used model for processes that vary continuously in space and/or time. Gaussian random fields are often used as building blocks in *Kriging* methods, a set of spatial prediction methods named after D.G. Krige. In Kriging approaches, the spatial process of interest can be modeled as a GRF, and a prediction at an unobserved location can be obtained through a weighted sum of observations from nearby locations (see e.g. Cressie (1993); Gelfand et al. (2010)).

While Gaussian random fields are used to model continuously indexed spatial data, Gaussian Markov random fields (GMRFs) are used to model spatial data that are discretely indexed, for example data that are connected to a grid or a graph. The spatial processes we model in this thesis are modeled as GRFs, but GMRF approximations are used to achieve computational benefits. Introductions to GRFs and GMRFs are given in the two next subsections.

1.2 Gaussian random fields (GRFs)

A Gaussian random field (GRF) is a probability model for a variable that varies over a continuous domain (see e.g. Stein (1999); Cressie (1993); Diggle and Ribeiro Jr (2007); Gelfand et al. (2010); Lieshout (2019)). Before defining a Gaussian random field, we define the well-known multivariate Gaussian distribution. A random vector $\boldsymbol{x} \in \mathcal{R}^n$ has a Gaussian distribution, denoted $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if its probability density function is given by

$$\pi(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-n/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right), \qquad (2)$$

where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is a $n \times n$ covariance matrix. The covariance matrix is symmetric positive semi-definite with element (i, j) given by $\mathrm{Cov}(x_i, x_j)$ where x_i and x_j are element i and j in $\boldsymbol{x}$. If two variables x_i and x_j are independent, $\mathrm{Cov}(x_i, x_j)$=0. For the multivariate normal distribution we also have that $\mathrm{Cov}(x_i, x_j) = 0$ implies independence.

We now define a GRF: A continuous field $\{x(\boldsymbol{u}); \boldsymbol{u} \in \mathcal{D}\}$ defined on a spatial domain $\mathcal{D} \in \mathcal{R}^d$ is a GRF if for any collection of locations $\boldsymbol{u}_1, ..., \boldsymbol{u}_n \in \mathcal{D}$ the vector $(x(\boldsymbol{u}_1), ..., x(\boldsymbol{u}_n))^{\mathrm{T}}$ follows a multivariate normal distribution,

i.e.

$$(x(\boldsymbol{u}_1), ..., x(\boldsymbol{u}_n))^{\mathrm{T}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}). \tag{3}$$

The elements in the covariance matrix can be constructed from a covariance function that defines the dependency between the locations in the spatial domain, and the covariance function is often characterized by two parameters; a spatial range ρ and a marginal standard deviation σ. The spatial range ρ defines how correlation decays with distance and is defined as the distance at which the correlation between two locations has dropped to almost zero. The marginal standard deviation σ quantifies spatial variability.

The mean $\boldsymbol{\mu}$ and the covariance model of a GRF can represent either stationary or non-stationary properties. A GRF is stationary if it has constant mean in the spatial domain and a covariance function that only depends on the distance between two locations $\boldsymbol{u}_i$ and $\boldsymbol{u}_j$, i.e.

$$\begin{aligned}
&\mathrm{E}\{x(\boldsymbol{u}_i)\} = \mu, \\
&\mathrm{Var}\{x(\boldsymbol{u}_i)\} = \sigma^2, \\
&\mathrm{Cov}\{x(\boldsymbol{u}_i), x(\boldsymbol{u}_j)\} = C(\boldsymbol{u}_i - \boldsymbol{u}_j),
\end{aligned}$$

for any i and j where $C(\cdot)$ is a covariance function. If we add the requirement that $C\{x(\boldsymbol{u}_i), x(\boldsymbol{u}_j)\} = C(||\boldsymbol{u}_i - \boldsymbol{u}_j||)$, where $||\boldsymbol{u}_j - \boldsymbol{u}_i||$ is the Euclidean distance between the locations $\boldsymbol{u}_j$ and $\boldsymbol{u}_i$, the field is also isotropic. On the contrary, if the mean and/or the covariance depend on location, the GRF is non-stationary.

There exist several covariance functions that can be used to construct stationary or non-stationary GRFs, with the Gaussian, exponential and Matérn covariance function being the most common. In the work presented in this thesis we use a stationary, isotropic Matérn covariance function to model the dependency structure of runoff. The stationary Matérn covariance function is given by

$$C(\boldsymbol{u}_i, \boldsymbol{u}_j) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa||\boldsymbol{u}_j - \boldsymbol{u}_i||)^{\nu} K_{\nu}(\kappa||\boldsymbol{u}_j - \boldsymbol{u}_i||), \tag{4}$$

where $||\boldsymbol{u}_j - \boldsymbol{u}_i||$ is the Euclidean distance between two locations $\boldsymbol{u}_i, \boldsymbol{u}_j \in \mathcal{R}^d$, K_{ν} is the modified Bessel function of second kind and order $\nu > 0$,

$\Gamma(\cdot)$ is the gamma function and σ^2 is the marginal variance (Guttorp and Gneiting, 2006). The parameter κ is a scale parameter, and it can be shown empirically that the spatial range can be expressed as $\rho = \sqrt{8\nu}/\kappa$, where ρ is defined as the distance where the spatial correlation between two locations has dropped to 0.1 (Lindgren et al., 2011).

The reasons for using the Matérn covariance function in our hydrological models are twofold: 1) The Matérn class of covariance functions has several useful properties and Stein (1999) strongly advices to use it. 2) Using the Matérn covariance function makes it possible to apply the SPDE approach to spatial modeling, which is based on approximating GRFs by GMRFs to achieve computational benefits. The SPDE approach is outlined in Section 1.5.2.

1.3 Gaussian Markov random fields (GMRFs)

While Gaussian random fields are connected to continuously indexed locations, Gaussian Markov random fields (GMRFs) are discretely indexed. GMRFs take spatial dependency into account through a neighborhood structure, and are often used for image analysis or to model areal level data. As GMRFs have *Markov properties*, computations with GMRF can be very efficient. We here define important concepts related to GMRFs. The definitions presented here are taken from Rue and Held (2005), and we refer to this book for further details.

Before presenting the formal definition of a GMRF, we define *conditional independence* and a *conditional independence graph*. Two variables x and y are called conditionally independent given z if and only if $\pi(x, y|z) = \pi(x|z)\pi(y|z)$. In the following, we denote this property as $x \perp y|z$. Furthermore, let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote a labeled graph with edges $\mathcal{E}$ and nodes $\mathcal{V} = \{1, ..., n\}$ that represent the elements of the random vector $\boldsymbol{x} = (x_1, ..., x_n)^{\mathrm{T}}$. Let $\mathcal{E}$ be such that there is no edge between node i and j if and only if $x_i \perp x_j|\boldsymbol{x}_{-ij}$, i.e. if and only if x_i is independent of x_j conditional on $\boldsymbol{x}_{-ij}$, where $\boldsymbol{x}_{-ij}$ denotes vector $\boldsymbol{x}$ without element x_i and x_j. The graph $\{\mathcal{V}, \mathcal{E}\}$ now defines a conditional independence graph.

The formal definition of a GMRF can now be formulated as follows (Defi-

nition 2.1 from Rue and Held (2005)):

A random vector $\boldsymbol{x} = (x_1, ..., x_n)^{\mathrm{T}} \in \mathcal{R}^n$ is called a GMRF with respect to a labeled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and symmetric positive definite precision matrix $\boldsymbol{Q}$, if and only if its density has the form

$$\pi(\boldsymbol{x}) = (2\pi)^{-n/2}|\boldsymbol{Q}|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{Q}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{5}$$

and

$$Q_{ij} \neq 0 \Longleftrightarrow \{i,j\} \in \mathcal{E} \text{ for all } i \neq j,$$

where $\boldsymbol{Q}$ is the inverse of the covariance matrix, $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$ and Q_{ij} denotes matrix element (i,j).

Comparing Equation (2) to Equation (5) we see that a GMRF is a GRF parametrized through the precision matrix instead of the covariance matrix. The GMRF parameterization is convenient because the precision matrix $\boldsymbol{Q}$ gives information about conditional independence properties of $\boldsymbol{x}$, i.e. the Markov properties. This can be seen if we compare the definition of the conditional independence graph with the GMRF definition. More specifically, $Q_{ij} = 0$ if and only if x_i is independent of x_j conditional on $\boldsymbol{x}_{-ij}$. The Markov properties of a GMRF can be summarized through the following theorem (Theorem 2.4 from Rue and Held (2005)):

Let $\boldsymbol{x}$ be a GMRF with respect to a labeled graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$. Then the following are equivalent.

1. The pairwise Markov property:

$$x_i \perp x_j | \boldsymbol{x}_{-ij} \quad \text{if } i \neq j \text{ and } \{i,j\} \notin \mathcal{E}.$$

2. The local Markov property:

$$x_i \perp \boldsymbol{x}_{-\{i,\mathrm{ne}(i)\}} | \boldsymbol{x}_{\mathrm{ne}(i)} \text{ for every } i \in \mathcal{V}.$$

3. The global Markov property:

$$\boldsymbol{x}_A \perp \boldsymbol{x}_B | \boldsymbol{x}_C$$

for all disjoint sets A, B and C where C separates A and B, and A and B are non-empty. Here, $\boldsymbol{x}_A = \{x_i : i \in A\}$ and similarly for $\boldsymbol{x}_B$ and

$\boldsymbol{x}_C$. Furthermore, ne(i) is all nodes in $\mathcal{G}$ with an edge to node i, i.e. the so-called neighbors of node i.

The above theorem implies that for a GMRF, the conditional distribution of x_i given $\boldsymbol{x}_{-i}$ can simply be expressed as the conditional distribution of x_i given its neighbors, i.e. as

$$\pi(x_i|\boldsymbol{x}_{-i}) = \pi(x_i|\boldsymbol{x}_{\text{ne}(i)}).$$

This property can also be read from the precision matrix $\boldsymbol{Q}$ of the GMRF, through its zero elements, or equivalently through the edges of the graph $\mathcal{G}$ that only exist between nodes that are conditionally dependent.

It is the many zero elements of $\boldsymbol{Q}$ (hence the GMRF's Markov properties) that give GMRFs computational benefits: There exist efficient algorithms for matrix operations on matrices with many zeros. For a GRF with a dense $n \times n$ covariance matrix the cost of matrix operations on $\boldsymbol{\Sigma}$ is $\mathcal{O}(n^3)$. The computational cost of matrix operations on the precision matrix of a GMRF depends on the GMRF itself and its conditional independence structure, but the computational cost is typically $\mathcal{O}(n^{3/2})$ for a two dimensional GMRF, a large reduction compared to the GRF. This is also the reason for using GMRFs in the models in thesis: Through the SPDE approach to spatial modeling, a GRF with a dense precision matrix can be approximated by a GMRF with a sparser precision matrix. See Section 1.5.2 for more on the SPDE approach, and Rue and Held (2005) for fast algorithms for GMRFs.

1.4 The Bayesian framework

The goal of any statistical analysis is to use observations to gain knowledge about a process of interest and/or to do predictions related to this process. There are two main approaches to statistics: The frequentist and the Bayesian. The frequentist approach can be considered as the more classical approach, while the Bayesian approach has gained more popularity in research communities later years. The reason for the increased popularity is related to recent developments in computer science and mathematics: Traditionally Bayesian inference and predictions were computationally expensive. Now there exist fast algorithms and servers

with more computational power, which have made "infeasible" Bayesian models computationally feasible. In this thesis, we operate in the Bayesian framework.

1.4.1 Bayesian statistics

In the traditional frequentist approach, a variable of interest $\boldsymbol{x}$ is considered fixed, but unknown. To make inference about the unknown quantity $\boldsymbol{x}$, relevant data $\boldsymbol{y}$ are collected. Based on knowledge about the process of interest, $\boldsymbol{x}$ and $\boldsymbol{y}$ are connected through an observation likelihood $\pi(\boldsymbol{y}|\boldsymbol{x})$, which expresses the probability of observing $\boldsymbol{y}$ for given values of the unknown variables $\boldsymbol{x}$. Next, by using e.g. maximum likelihood procedures $\boldsymbol{x}$ can be estimated based on $\boldsymbol{y}$,

What is different in the Bayesian framework, is that the variable of interest $\boldsymbol{x}$ is not longer considered as fixed, but as a quantity whose variation can be expressed through a probability distribution (see e.g. Casella and Berger (1990); Berger (1985)). Before performing the statistical analysis, this probability is expressed through a so-called prior distribution $\pi(\boldsymbol{x})$. This can be constructed based on expert knowledge about the variable(s) of interest. The goal of the Bayesian analysis, is to update the prior distribution based on data $\boldsymbol{y}$. By using Bayes' theorem

$$\pi(\boldsymbol{x}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{x})\pi(\boldsymbol{y}|\boldsymbol{x})}{\pi(\boldsymbol{y})} \propto \pi(\boldsymbol{x})\pi(\boldsymbol{y}|\boldsymbol{x}), \tag{6}$$

the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ is obtained. Here, $\propto$ means proportional to. The posterior distribution expresses the degree of belief in each possible value of $\boldsymbol{x}$ given the data. To obtain the marginal distribution for any element $x_i \in \boldsymbol{x}$, the remaining variables are integrated out, and a prediction of x_i can be summarized through e.g. the mean, median or the mode of the posterior distribution $\pi(x_i|\boldsymbol{y})$.

Three appealing properties of the Bayesian approach is that 1) we obtain a full predictive distribution for the variable of interest $\boldsymbol{x}$, 2) it provides a convenient framework for hierarchical modeling which is further described in Section 1.4.2 and 3) the possibility to incorporate expert opinions into the modeling through the prior distributions. However, the latter requires

knowledge about how subjective prior beliefs should be translated into mathematically formulated prior distributions. The Bayesian approach can also produce posterior distributions that are heavily influenced by the priors. Another challenge is the computational complexity associated with many Bayesian models. We discuss prior modeling and the computational challenges in Bayesian statistics in Section 1.4.3 and Section 1.5 respectively.

1.4.2 Hierarchical modeling and latent Gaussian models (LGMs)

The class of hierarchical models represents a popular and flexible modeling framework that is used both within the Bayesian and the frequentist approach to statistics. In a hierarchical model, the statistical model is formulated in multiple levels, i.e. in a hierarchical form (see e.g. Gelman et al. (2004); Gelman and Hill (2007); Banerjee et al. (2004)). Each level is given a distribution and can be considered as a sub-block of the full model. By multiplying the distributions of all sub-blocks together, the joint distriution of the model quantities of interest is obtained. This way complicated processes can be modeled as a sequence of simpler sub-models.

A hierarchical model is often characterized by three stages. The first stage is the *data model*: Data $\boldsymbol{y} = (y_1, .., y_n)$ are observed with conditionally independent likelihood $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$ given a latent (unobserved) process $\boldsymbol{x}$ and some model parameters $\boldsymbol{\theta}$. The second stage is the *process model* or the *latent model*. This consists of the probabilistic model for the latent, underlying process $\boldsymbol{x}$ given the parameters: $\pi(\boldsymbol{x}|\boldsymbol{\theta})$. The third stage is the *parameter model* $\pi(\boldsymbol{\theta})$, a model for the parameters. In a Bayesian setting, the parameter model $\pi(\boldsymbol{\theta})$ and the process model $\pi(\boldsymbol{x}|\boldsymbol{\theta})$ are given by the prior distributions of $\boldsymbol{x}$ and $\boldsymbol{\theta}$.

To make inference about the unknown parameters $\boldsymbol{\theta}$ and variables $\boldsymbol{x}$ in a hierarchical model, Bayes' rule is used:

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{y})} \propto \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{7}$$

Through marginalization, the distributions $\pi(\boldsymbol{x}|\boldsymbol{y})$ and $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ are obtained. Within the Bayesian framework, these are the posterior distributions of the quantities of interest.

If using the frequentist approach for computing $\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y})$, or an empirical Bayesian approach, the parameters $\boldsymbol{\theta}$ are typically estimated first by using e.g. maximum likelihood procedures. Next, $\boldsymbol{\theta}$ is treated as fixed in Equation (7), with value equal to the estimated value $\hat{\boldsymbol{\theta}}$. This means that the uncertainty of $\boldsymbol{\theta}$ is not taken into account in the assessment of $\pi(\boldsymbol{x}|\boldsymbol{y})$. This is different from the fully Bayesian approach, where both $\boldsymbol{\theta}$ and $\boldsymbol{x}$ are associated with prior and posterior distributions. The full uncertainty specification for the parameters provided by the Bayesian approach, is one of the reasons why the Bayesian framework is particularly suitable for hierarchical statistical modeling.

In this thesis, (fully) Bayesian hierarchical models are used to build hydrological models, and we use a specific class of hierarchical models called latent Gaussian models (LGMs). An LGM is a natural modeling choice when considering spatial processes that can be modeled as GRFs. Another main motivation for using LGMs is that it makes it possible to use INLA to make faster inference and predictions (Rue et al., 2009).

The INLA approach supports LGMs where the mean μ_i of the observation y_i is linked to a linear predictor through a link function $g(\cdot)$ on the form:

$$g(\mu_i) = \eta_i = \beta_0 + \sum_{k=1}^{n_\beta} \beta_k z_{ki} + \sum_{j=1}^{n_f} f^{(j)}(u_{ij}) + \epsilon_i, \tag{8}$$

where $\boldsymbol{x} = (\beta_0, \{\beta_k\}, \{f^{(j)}(\cdot)\}, \{\eta_i\})$ defines the latent field of the hierarchical model. Here, β_0 is an intercept, $\{f^{(j)}(\cdot)\}$ are functions of $\{u_{ij}\}$ that can be used to model e.g. spatial or temporal dependency structures, $\{\beta_k\}$ are the linear effects of some covariates $\{z_{ki}\}$ and $\{\epsilon_i\}$ are independent and identically distributed error terms. All components in $\boldsymbol{x}$ must be given Gaussian priors in order to obtain an LGM. Based on the latent model defined in Equation (8), the observation likelihood of the LGM can be written as $\pi(y_i|\boldsymbol{x}, \boldsymbol{\theta}) \sim \pi(y_i|\eta_i, \boldsymbol{\theta})$. The third stage of the LGM is given by the prior of the model parameters $\pi(\boldsymbol{\theta})$ as before. The additive structure of several components, introduced by the linear predictor in Equation (8), allows for

a large range of models with different dependency structures. The INLA approach is outlined in Section 1.5.1.

1.4.3 Knowledge-based, interpretable priors

One of the challenges in Bayesian modeling, is to specify reasonable prior distributions. The prior distributions should reflect what the researcher knows about the parameters of interest. In a hierarchical Bayesian framework, it is often relatively simple to select a suitable prior $\pi(\boldsymbol{x}|\boldsymbol{\theta})$ for the latent process. Constructing priors for the model parameters $\boldsymbol{\theta}$ however, is typically less intuitive.

A prior for $\boldsymbol{\theta}$ can be *informative*, *vague* or *non-informative* (see e.g. Chapter 5 in Lesaffre and Lawson (2012) or Chapter 3 in Blangiardo and Cameletti (2015)). Non-informative priors are typically used if the researcher has limited information or intuition about the parameters of interest and/or want the data to determine. A non-informative prior can be constructed by assigning an equal probability to all possible values of the parameter of interest by using a uniform distribution. However, a problem with the uniform distribution is that it is not invariant to transformations. As an alternative to the uniform distribution, several non-informative priors that are transformation invariant have been suggested. Two popular choices are Jeffreys' prior (Jeffreys, 1946) and the *reference prior* suggested by Bernardo (1979). While Jeffreys' prior is a function of the Fisher information, and hence also the observation likelihood, the reference prior is constructed by maximizing the Kullback–Leibler distance between the prior and the posterior. A drawback of Jeffreys' and Bernardo's priors however, is that they often are improper, i.e. they don't integrate to 1. This is problematic because it for certain models also can result in improper posterior distributions. To avoid improper posteriors, so-called *vague* priors can be used as approximations for non-informative priors. An example of a vague prior can e.g. be a Gaussian prior distribution for a regression coefficient with (very) large prior variance. However, despite of being constructed to be vague, these priors often lead to quite informative priors (Gelman, 2006; Fong et al., 2009).

When the researcher has information about the process of interest, *infor-*

mative priors can be used to incorporate his or her knowledge into the Bayesian model. The information can be based on e.g. results from previous experiments, knowledge about the measurement uncertainty of the variable(s) of interest, or if it is a spatial model, knowledge about the study area. An informative prior for a parameter that has been estimated to be in the interval [0,10] for similar experiments, can for example simply be a Gaussian distribution centered around 5.

Specification of priors is a major challenge in Bayesian statistics and a source of criticism from frequentists. In many applications priors are set in an ad-hoc manner, and/or default priors in computer software such as `r-inla` are used (Rue et al., 2009). This can lead to overfitting, prior sensitive results and it often makes prior assumptions unavailable for other researchers or users. Later years, research has been done to make prior specification easier. One of the contributions here is the PC prior suggested by Simpson et al. (2017). The penalized complexity (PC) prior is based on penalizing the increased complexity of a model induced by deviating from a simpler base model. This is done by using a function of the Kullback-Leibler divergence to measure the increased complexity between a base model and the proposed, more flexible model. The goal here is to control against overfitting. Another benefit of the PC prior is that it is relatively easy to interpret for the user. The prior can be specified through a probability $\alpha \in (0, 1)$ and a quantile $U > 0$ on the form

$$Prob(Q(\xi) > U) = \alpha \tag{9}$$

where $Q(\cdot)$ is an interpretable transformation of the parameter of interest ξ. To demonstrate how this is done, we use the PC prior for a Gaussian effect $\mathcal{N}(0, \tau^{-1})$ as an example. This is given by the density

$$\pi(\tau) = \frac{\lambda}{2}\tau^{-3/2}\exp(-\lambda\tau^{-1/2}), \qquad \tau > 0, \quad \lambda > 0, \tag{10}$$

where λ is a parameter that determines the penalty of deviating from the base model. However, the parameter λ, and hence also the PC prior for τ, can simply be specified through $\text{Prob}(1/\sqrt{\tau} > U) = \alpha$, where $1/\sqrt{\tau}$ is the standard deviation of the Gaussian effect and where $\lambda = -\ln(\alpha)/u$. Hence, the user only needs to have a intuition about the standard deviation of the Gaussian effect $N(0, \tau^{-1})$ and can formulate his/her thoughts through α

and U. In this thesis, PC priors are used for the variance or precision parameters in our models.

In our work, we also apply the joint PC prior suggested by Fuglstad et al. (2019b) to specify priors for the range ρ and the marginal standard deviation σ of a GRF. This PC prior can be specified through the quantiles U_ρ, U_σ and probabilities α_ρ, α_σ as

$$\text{Prob}(\rho < U_\rho) = \alpha_\rho; \qquad \text{Prob}(\sigma > U_\sigma) = \alpha_\sigma,$$

where u_ρ, U_σ, α_ρ and α_σ are quantiles and probabilities that must be specified by the modeler.

The fact that the PC priors can be specified through quantiles and probabilities makes prior specification, and hence also the Bayesian framework, more available for researchers that don't have a strong mathematical background. Prior assumptions also become easier to communicate to users (e.g. hydrologists) and/or to criticize. In the field of making prior specification simpler, more interesting work is going on. One example is the hierarchical decomposition (HD) prior that was recently developed by Fuglstad et al. (2019a). The HD prior framework can be used to construct priors for the total variance in an additive model by decomposing the total variance through a tree structure of the underlying, individual model components.

1.5 Computational challenges in Bayesian statistics and in spatial modeling

The goal in Bayesian analysis is to determine the (marginal) posterior distributions of the latent field $\pi(\boldsymbol{x}|\boldsymbol{y})$ and/or the parameters $\pi(\boldsymbol{\theta}|\boldsymbol{y})$. To achieve this analytically, an evaluation of $\pi(\boldsymbol{y})$ is needed (Equation (6)). However, $\pi(\boldsymbol{y})$ is often not analytically tractable: To obtain $\pi(\boldsymbol{y})$ one has to integrate the joint distribution $\pi(\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\theta})$ over $\boldsymbol{\theta}$ and $\boldsymbol{x}$, where the dimension of $\boldsymbol{\theta}$ and $\boldsymbol{x}$ can be large. For this reason, Bayesian analysis on complex models was not available until the 1990s when computational sampling algorithms were developed. Among these we find the well-known Gibbs sampling procedure and other Markov chain Monte Carlo (MCMC) approaches (Gelfand and Smith, 1990; Casella and George, 1992). These

have gained popularity through user friendly computer programmes like BUGS (Spiegelhalter et al., 1999), JAGS (Plummer, 2003) and `Rstan` (Carpenter et al., 2017). For many applications, sampling-based MCMC methods still represent the best approach available for performing Bayesian analysis. However, if the Bayesian model is complex with large dependencies between the components and/or if the dimension of $\boldsymbol{x}$ and $\boldsymbol{\theta}$ is large, MCMC can be slow and computationally unfeasible within reasonable time.

There are also computational challenges related to fitting geostatistical models involving GRFs. The computational complexity is connected to the matrix operations that are required on the covariance matrix $\boldsymbol{\Sigma}$ to do predictions and inference. Inverting a dense $n \times n$ covariance matrix $\boldsymbol{\Sigma}$ has a computationally complexity of $\mathcal{O}(n^3)$ when using standard inversion algorithms. In spatial analysis, n is the number of target locations (observation and prediction locations), and this number is often large.

In this thesis, we mostly consider Bayesian hierarchical models involving not only one, but two GRFs. The number of target locations n is also large, considering that we for example are interested in making a runoff map for the whole country of Norway for up to 30 different years. Consequently, our models are computationally demanding. These challenges would have made the suggested runoff models computationally unfeasible only a decade ago, at least for many operational applications and for cross-validation assessments. However, relatively recent developments in statistics enable (approximate) inference and predictions on more complex Bayesian and/or spatial models. We here particularly refer to the INLA approach developed by Rue et al. (2009), and to the SPDE approach to spatial statistics developed by Lindgren et al. (2011). While INLA is used to make Bayesian inference, the SPDE approach is used to overcome the challenges related to inversion and multiplication of dense covariance matrices. In the following two subsections, we give an overview of INLA and the SPDE approach.

1.5.1 Integrated nested Laplace approximations (INLA)

INLA is a tool for making approximate Bayesian inference on LGMs and was suggested by Rue et al. (2009). It can be used for LGMs with linear

predictors on the form expressed in Equation (8) if each component of the latent field $\boldsymbol{x}$ is given a Gaussian prior distribution.

The goal of any Bayesian analysis is to compute the posterior marginal distributions

$$\pi(x_i|\boldsymbol{y}) = \int \pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}, \quad i = 1, ..., n_x \tag{11}$$

$$\pi(\theta_k|\boldsymbol{y}) = \int \pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}_{-k}, \quad k = 1, ...n_\theta, \tag{12}$$

where $\boldsymbol{\theta}$ is the parameter vector from Section 1.4.2 and $\boldsymbol{x}$ is the latent field. Here, x_i denotes element i of vector $\boldsymbol{x}$ and $\boldsymbol{x}_{-i}$ denotes vector $\boldsymbol{x}$ without element x_i. Similarly, θ_k denotes element k of vector $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{-k}$ denotes the parameter vector without element θ_k.

To evaluate the two densities above, evaluations of $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ and $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ are needed. The INLA approach computes these densities by using Laplace approximations. A Laplace approximation can be used to approximate any density $g(x)$ with a normal distribution (see Tierney and Kadane (1986)). In INLA this is used to express the posterior density $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ from Equation (11) and (12) as

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})} \approx \left.\frac{\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\tilde{\pi}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}\right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}, \tag{13}$$

where $\tilde{\pi}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ is the Laplace approximation of the density $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$, and $\boldsymbol{x}^*(\boldsymbol{\theta})$ is its mode for a given $\boldsymbol{\theta}$. The INLA methodology finds suitable evaluation points for $\boldsymbol{\theta}$ by using a numerical optimization algorithm. See Rue et al. (2009) for details.

Next, $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ in Equation (11) must be evaluated to make inference, for for $i = 1, .., n_x$. This is more difficult and computationally expensive because the dimension of $\boldsymbol{x}$ typically is larger than the dimension of $\boldsymbol{\theta}$. To approximate $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$ three approaches are suggested by Rue et al. (2009): The first approach is to use a Gaussian approximation for $\pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})$. This is already available through $\tilde{\pi}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ from Equation (13). However, this approximation often leads to inaccurate results. A second approach is to

use

$$\pi(x_i|\boldsymbol{\theta},\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})}{\pi(\boldsymbol{x}_{-i}|x_i,\boldsymbol{\theta},\boldsymbol{y})} \approx \frac{\pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y})}{\tilde{\pi}(\boldsymbol{x}_{-i}|x_i,\boldsymbol{\theta},\boldsymbol{y})}\bigg|_{\boldsymbol{x}_{-i}=\boldsymbol{x}^*_{-i}(x_i,\boldsymbol{\theta})} =: \tilde{\pi}(x_i|\boldsymbol{\theta},\boldsymbol{y}), \tag{14}$$

i.e. the density $\pi(\boldsymbol{x}_{-i}|x_i,\boldsymbol{\theta},\boldsymbol{y})$ is approximated by a Laplace approximation and evaluated at its mode $\boldsymbol{x}^*_{-i}(x_i,\boldsymbol{\theta})$. This approximation is good, but the drawback is that it requires an evaluation of $\tilde{\pi}(\boldsymbol{x}_{-i}|x_i,\boldsymbol{\theta},\boldsymbol{y})$ for all values of $\boldsymbol{\theta}$ and $\boldsymbol{x}$. This is computationally expensive. Finally, the third approach is to use the *simplified Laplace approximation*, which is based on a Taylor's series expansion of the Laplace approximation $\tilde{\pi}(x_i|\boldsymbol{\theta},\boldsymbol{y})$ in Equation (14). The latter is the default approach in the `R` package `r-inla` and is quite fast and accurate for most model types.

After approximating $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ and $\pi(x_i|\boldsymbol{\theta},\boldsymbol{y})$, the integrals in Equation (11) and (12) must be evaluated. This is done based on numerical integration schemes, and we again refer to Rue et al. (2009) for more.

In the calculations described above, INLA relies on using the fast algorithms for matrix operations on GMRF's from Rue and Held (2005). Hence, in order to obtain fast inference and predictions with INLA, the latent field of the LGM $\boldsymbol{x}$ should be a GMRF with sparse precision matrix. Another requirement for achieving fast predictions is that the number of so-called hyperparameters $\boldsymbol{\theta}$ in the model should not be too large, say smaller than 15. The reason is that we in Equation (11) and (12) need to integrate over the hyperparameter space in order to obtain the results, and this is computationally infeasible if the dimension of $\boldsymbol{\theta}$ is large.

Apart from this, INLA is in general accurate and fast for most LGMs (Rue et al., 2009; Martino et al., 2011; Eidsvik et al., 2012; Huang et al., 2017), and has been successfully used within several fields of science (see e.g. Khan and Warner (2018); Opitz et al. (2018); Yuan et al. (2017); Guillot et al. (2014); Ingebrigtsen et al. (2014); Myrvoll-Nilsen et al. (2020); Jullum et al. (2020)). However, INLA has faced problems for some (more extreme) models with binomial or Poisson data (Fong et al., 2009; Ferkingstad and Rue, 2015). In this thesis, we use Gaussian likelihoods which implies that the Laplace approximation is exact. The accuracy of INLA is then only determined by the accuracy of the numerical integration schemes used

to evaluate Equation (11) and (12), and we can expect INLA to provide reliable results.

The INLA methodology is available through the R package `r-inla`, and we refer to Blangiardo and Cameletti (2015), Krainski et al. (2018) and the web page `www.r-inla.org` for `r-inla` tutorials and examples.

1.5.2 The SPDE approach to spatial modeling

The INLA approach to Bayesian inference requires that the precision matrix of the latent Gaussian field is sparse in order to work fast. This is typically not the case for spatial models that involves GRFs. However, the SPDE approach to spatial modeling suggested by Lindgren et al. (2011), makes it possible to approximate GRFs by GMRFs with a sparser precision matrices. The approach is based on the following stochastic partial differential equation (SPDE)

$$(\kappa^2 - \Delta)^{\alpha/2}\tau x(\boldsymbol{u}) = \mathcal{W}(\boldsymbol{u}), \quad \boldsymbol{u} \in \mathcal{R}^d, \quad \kappa > 0, \tag{15}$$

whose exact and stationary solution $x(\boldsymbol{u})$ is a GRF with Matérn covariance matrix (Equation (4)) as shown by Whittle (1954, 1963). In Equation (15), $\mathcal{W}(\cdot)$ is spatial Gaussian white noise, Δ is the Laplacian, α is a smoothness parameter, κ is the scale parameter from the Matérn covariance function from Equation (4), d is the dimension of the spatial domain and τ is a parameter controlling the variance. The parameters τ, κ, α and d of the above SPDE are connected to the parameters of the Matérn covariance function in Equation (4) through

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}\kappa^{2\nu}\tau^2}; \qquad \nu = \alpha - d/2, \tag{16}$$

where σ^2 is the marginal variance of the Matérn covariance function, ν is its smoothness parameter and $\Gamma(\cdot)$ is the gamma function.

In Lindgren et al. (2011) the SPDE in Equation (15) is solved numerically by using the finite element method (FEM) (see e.g. Brenner and Scott (2008). Through FEM, an approximation of the solution can be written as

$$x(\boldsymbol{u}) = \sum_{k=1}^{m} \phi_k(\boldsymbol{u}) w_k. \tag{17}$$

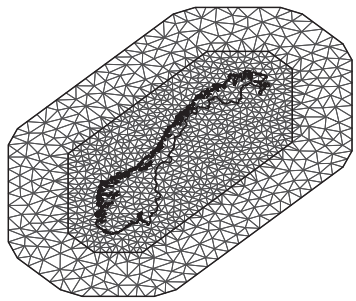


Figure 1: The finite element method is used to find an approximate solution of the SPDE in Equation (15). The solution is expressed through a basis function representation defined on a triangular mesh. Above we see an example mesh for Norway. The number of mesh nodes m determines the computational complexity of performing inference and predictions on the approximative GMRF solution of the SPDE. Also note that the above example mesh has an outer layer to handle boundary effects.

This is a basis function representation of the GRF $x(\boldsymbol{u})$ defined on a triangulation of the domain $\mathcal{D}$, where m is the number of vertices in the triangulation, ϕ_k for $k = 1, .., m$ are basis functions and w_k for $k = 1, ..m$ are zero mean Gaussian distributed weights. The weights are approximations of the GRF $x(\boldsymbol{u})$ at the mesh nodes, and the basis functions transfer the approximation from the mesh nodes to other locations of interest $\boldsymbol{u}$. An example triangulation mesh is shown in Figure 1. The basis functions ϕ_k are chosen such that they are piecewise linear in each mesh triangle and have a local support. That means that $\phi_k = 1$ at vertex k and 0 at all other vertices. This ensures that the approximation in Equation (17) has Markov properties and forms a GMRF. Assuming Neumann boundary conditions, and $\alpha = 2$ in Equation (15), the precision matrix for the Gaussian weights (w_1,...,w_m) is given by

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

where $\boldsymbol{C}$ is a diagonal matrix with elements given by $C_{ii} = \int \phi_i(\boldsymbol{u})d\boldsymbol{u}$ and $\boldsymbol{G}$ is a sparse matrix with elements $G_{ij} = \int \Delta\phi_i(\boldsymbol{u})\Delta\phi_j(\boldsymbol{u})d\boldsymbol{u}$ with Δ denoting the gradient. The precision matrix $\boldsymbol{Q}$ is sparse, hence $w_1, .., w_m$ and also the approximation in Equation (17) forms a GMRF.

The possibility of approximating GRFs by GMRFs through the SPDE approach has made many complex Bayesian spatial models computationally feasible. The computational cost of matrix operations on the GMRF approximation from Equation (17) is $\mathcal{O}(m^{3/2})$, where m is the number of nodes in the triangulation mesh. Recall that for a GRF with dense covariance matrix, the computational cost of matrix operations is $\mathcal{O}(n^3)$, where n is the number of target locations. Hence, it is the mesh size determines the computational complexity of the problem in the SPDE approach, and not the number of observation and prediction locations. In a typical setting can m be smaller than n and the SPDE approach still gives accurate predictions. However, if the mesh is too coarse it of course affects the accuracy of the results: The edges in the mesh should be defined such that they are short enough to capture the spatial variability in the study area. The mesh should also be extended outside the target area, as in Figure 1, such that it is wide enough to handle the boundary effects that can occur from solving the SPDE numerically.

Since the SPDE approach's introduction it has been used in several applications (see e.g. Wakefield et al. (2017); Bakka et al. (2018); Moraga et al. (2017); Jullum et al. (2020); Lenzi et al. (2018)), and it is available in the `R` package `r-inla` together with the INLA approach. One of the benefits of the SPDE approach, in addition to the computational benefits, is that it makes it simple to introduce non-stationary models by allowing the SPDE parameters τ and κ to vary in space e.g. as functions of explanatory variables (Ingebrigtsen et al., 2014, 2015; Bakka et al., 2019). As non-stationarity can be expressed through the SPDE parameters, we don't need to explicitly construct a valid, positive definite covariance matrix. This is automatically ensured.

The SPDE approach is developed to handle the computational complexity of geostatistical models. There also exist other approaches that try to handle computational expensive spatial models, like fixed rank Kriging (Cressie and Johannesson, 2008), predictive process models (Banerjee et al., 2008) and covariance tapering (Furrer et al., 2005). Presenting these methods are outside the scope of this thesis, but we include them here as a reference.

2 Contributions towards annual runoff interpolation and the prediction in ungauged basins problem

This thesis contains three papers, Paper A-C, where all of them are motivated by the prediction in ungauged basins problem in hydrology, i.e. the problem of predicting runoff in areas that lack runoff observations. In the following subsections we explain some of the challenges that exist in runoff modeling, and how other researches have addressed them. We also describe how our work has contributed towards the field, and how the statistical tools described in Section 1 are used to build new hydrological models.

2.1 Areal referenced runoff models (Paper A-C)

Every point in the landscape contributes to runoff generation, and we can think of runoff as a continuous process in space. However, when measuring runoff, it is done by measuring the volume of water per unit time that flows through a cross-section of a stream or river. These observations are typically given in mm per time unit or in m^3/s and contain information about the *average* runoff in the drainage area of the target river, i.e. what we call a *catchment* area. Runoff observations are therefore areal referenced. Furthermore, catchments are organized in a nested structure such that the runoff observations often come from catchments that overlap. This has two major consequences in a modeling setting:
1) The nestedness of catchments affects the dependency structure of runoff.
2) Mass preservation laws matter. It is impossible to have a larger water volume in a small subcatchment than in a larger overlapping catchment.

The above challenges are well known by hydrologists and are typically accounted for in commonly used hydrological process-based models like the HBV model (Sveriges Meteorologiska och Hydrologiska Institut, 1992) and WASMOD (Widén-Nilsson et al., 2007). However, for most statistical runoff models the situation is different. Methods like multiple linear regression (Engeland et al., 2020; Merz and Blöschl, 2005) don't ensure consistent predictions over nested catchments, and neither does the ROI approaches

(Zrinji and Burn, 1994; Merz and Blöschl, 2005). Within the geostatistical framework, many of the models consider runoff observations as point referenced instead of areal referenced (see e.g. Merz and Blöschl (2005); Skøien et al. (2003); Adamowski and Bocci (2001)). Here, the catchment centroids or the locations of the stream gauges are used as the observation locations, and the modeler risks a violation of conservation laws and a poor uncertainty model. However, there exist a few geostatistical models that take the nested structure of catchments into account in the runoff interpolation. Central examples are the models in Skøien et al. (2006); Sauquet et al. (2000); Gottschalk (1993a,b). The Top-Kriging approach suggested by Skøien et al. (2006) has given particularly good results (Viglione et al., 2013; Blöschl et al., 2013), and is considered as one of the best interpolation methods for hydrological data.

As the name suggests, Top-Kriging is a Kriging approach. In Kriging approaches, the spatial variable of interest $x(\boldsymbol{u})$ is often modeled as a GRF. A prediction of the target variable $x(\boldsymbol{u}_0)$ at an unobserved location $\boldsymbol{u}_0 \in \mathcal{R}^2$ is given by a weighted sum of the observed values at $\boldsymbol{u}_1, ..., \boldsymbol{u}_n$, i.e. as

$$\hat{x}(\boldsymbol{u}_0) = \sum_{i=1}^{n} \lambda_i x(\boldsymbol{u}_i), \tag{18}$$

where the λ_i's are interpolation weights that must be determined and the $x(\boldsymbol{u}_i)$'s are observations from nearby locations. The interpolation weights are estimated by requiring that $\hat{x}(\boldsymbol{u}_0)$ is the Best Linear Unbiased Estimator (BLUE) of $x(\boldsymbol{u}_0)$, which means that $\hat{x}(\boldsymbol{u}_0)$ is estimated by finding the weights that both minimize the mean squared error and give zero mean expected error (Cressie, 1993). This requires evaluations of the covariance function of the involved GRF. What makes Top-Kriging different from other hydrological Kriging approaches in this context, is that Top-Kriging accounts for the nested structure of catchments in the covariance (variogram) calculations by interpreting the runoff observations as areal referenced instead of point referenced. The catchment areas are discretized, and the distance between two catchments is measured by calculating pairwise distances between the grid nodes contained by the target catchments' discretizations. By this, a subcatchment gets a higher Kriging weight λ_i than a nearby, non-overlapping catchment.

Apart from the geostatistical models in Skøien et al. (2006); Sauquet et al. (2000); Gottschalk (1993a,b), there are few established geostatistical models in hydrology that treat runoff as areal referenced, and there is more work that can be done within this field. One of the reasons for the little amount of research, might be that an areal referenced model often represent a substantial increase in computational complexity compared to a point referenced model. However, such models are now more available through tools as INLA and the SPDE approach to spatial modeling.

In Paper A - C we suggest Bayesian interpolation models for annual runoff that treat runoff as areal referenced. In our models, annual runoff $q(\boldsymbol{u})$ is modeled as a continuous spatial process $\{q(\boldsymbol{u}) : \boldsymbol{u} \in \mathcal{D}\}$ that can be defined for any point in the landscape $\mathcal{D} \in \mathcal{R}^2$. However, runoff is observed at catchment level, and we model the true runoff generated inside a catchment $\mathcal{A}$ as

$$Q(\mathcal{A}) = \frac{1}{|\mathcal{A}|} \int_{\boldsymbol{u} \in \mathcal{A}} q(\boldsymbol{u}) d\boldsymbol{u} \approx \frac{1}{n_{\mathcal{A}}} \sum_{\boldsymbol{u} \in \mathcal{L}_{\mathcal{A}}} q(\boldsymbol{u}), \tag{19}$$

i.e. as the average point runoff $q(\boldsymbol{u})$ over the catchment, where $|\mathcal{A}|$ is the catchment area. The model is given conditional on the model parameters. To make the model computationally feasible, the catchments are discretized, and the integral in Equation (19) is replaced by a sum over the discretization grid $\mathcal{L}_{\mathcal{A}}$, where $n_{\mathcal{A}}$ is the number of grid nodes in the discretization of catchment $\mathcal{A}$. The spatial models we use for $q(\boldsymbol{u})$ are specified later, in Section 2.3 and Section 2.4, but for now we only assume that $q(\boldsymbol{u})$ is a stationary, continuous spatial model that does not depend on any explanatory variables.

The true runoff from Equation (19) is not observed directly, but through noisy streamflow observations. We use the following observation likelihood for runoff, that is given conditional on $Q(\mathcal{A})$:

$$\begin{aligned} z_i &= Q(\mathcal{A}_i) + \epsilon_i^z \\ \pi(z_i | Q(\mathcal{A}_i), \sigma_z) &\sim \mathcal{N}(Q(\mathcal{A}_i), s_i^z \sigma_z^2). \end{aligned} \tag{20}$$

Here, z_i is the observed runoff in catchment $\mathcal{A}_i$, ϵ_i^z are independent and identically distributed as $\mathcal{N}(0, s_i^z \sigma_z^2)$, σ_z^2 is a variance parameter and s_i^z is a fixed, predetermined scale that is specific for catchment $\mathcal{A}_i$ allowing for differences in measurement uncertainty over different catchments.

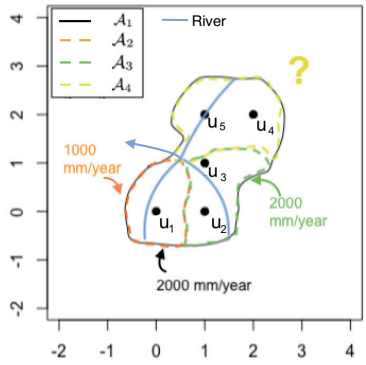


Figure 2: Conceptual figure of four discretized catchments where each grid node $\boldsymbol{u}_1, ...\boldsymbol{u}_5$ represents one areal unit. Assume that we have the following observations of annual runoff from $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$: $Q(\mathcal{A}_1)$=2000 mm/year, Q($\mathcal{A}_2$)=1000 mm/year and $\mathcal{A}_3$=2000 mm/year. In order to fulfill the water balance constraints in Equation (19), imposed by the likelihood in Equation (20), the predicted mean annual runoff in the ungauged catchment ($\mathcal{A}_4$) must be around 2500 mm/year if we assume a low observation uncertainty. Mark that we use a very coarse grid here to make the example more intuitive.

Compared to Top-Kriging, the above model formulation gives a similar covariance model. It decreases the posterior uncertainty inside the areas of the observed catchments, not only at certain gauging points which is what a point referenced model would do. It also weights an observation from a subcatchment more than an observation from a nearby non-overlapping catchment when making predictions, similar to Top-Kriging. On the other hand, our model formulation is different from Top-Kriging by that we impose soft constraints on the predicted runoff within the areas of the observed catchments. This is done through the process model in Equation (19) and the observation likelihood in Equation (20). The equations in (19) and (20) actually correspond to water balance constraints.

The consequence of imposing water balance constraints in our model, is that the model is influenced to follow basic preservation laws. This gives consistent runoff predictions (posterior mean) over nested catchments. An additional consequence of imposing water balance constraints, is that the proposed model is able to predict larger values than any of the observed val-

ues without using explanatory variables. One simplified example is shown in Figure 2. Here, we assume that we have three observations of runoff on a simplified grid: 2000 mm/year in Catchment $\mathcal{A}_1$, 1000 mm/year in Catchment $\mathcal{A}_2$ and 2000 mm/year in Catchment $\mathcal{A}_3$. Using equations (19) and (20), we can specify three soft constraints, one per observed catchment. It can be shown (and also understood intuitively from Figure 2) that in order to fulfill the constraints, the predicted runoff in the unobserved catchment $\mathcal{A}_4$ must be:

$$\hat{Q}(\mathcal{A}_4) = \frac{q(\boldsymbol{u}_4) + q(\boldsymbol{u}_5)}{2} + \text{uncertainty} = 2500 \text{ mm/year} + \text{uncertainty} \quad (21)$$

where the *uncertainty* term is determined by the observation uncertainty for a fixed set of marginal variances and ranges. Hence, we can obtain a prediction that is larger than any of the observed values if the observation uncertainty is sufficiently small. This property is not held by most other geostatistical methods that only rely on spatial smoothing. In Paper B we also explain how we construct informative priors with narrow credible intervals on the measurement uncertainty $s_i^z \sigma_z^2$ in order to make the water balance constraints stronger.

Top-Kriging does not constrain the runoff over nested catchments as in the example above. Hence, its predictions can more easily violate the water balance. This can either represent a benefit or a drawback relative to our method, depending on what kind of variable we want to model: Not all hydrological variables are mass conserved. Another existing hydrological model that considers water balance constraints in the interpolation, is the geostatistical model for annual runoff from Sauquet et al. (2000). This is a Kriging approach where mass balance constraints are used as additional constraints in the Kriging system of equations.

2.2 Combining point and areal referenced data (Paper A)

Using precipitation and evaporation data as observed values when performing runoff interpolation is something that in general has been avoided by Top-Kriging and other interpolation methods used in hydrology. The reason is expressed in Viglione et al. (2013) as follows:

"The main advantage of statistical methods for estimating runoff in ungauged basins is that they avoid the use of uncertain input variables such as precipitation and potential evaporation".
Precipitation data are uncertain and biased due to *undercatch* (Wolff et al., 2015), while evaporation is seldom observed directly, but provided by algorithms that use meteorological observations from satellites as input (see e.g. Mu et al.; Zhang et al. (2009)). In spite of the uncertainties linked to these observation types, evaporation and precipitation are the main drivers behind runoff as we can see from the basic water balance equation in (1). Hence, it is reasonable to assume that precipitation and evaporation data can provide useful information about runoff, particularly in areas of the world where there are few streamflow observations of runoff available. However this requires that the precipitation and evaporation data are linked to runoff in a clever way. A geostatistical model for runoff that allows evaporation and precipitation in the observations likelihood also has to take into account that precipitation and evaporation data are observed at point locations, while streamflow observations are areal referenced: The observations of different spatial support should be connected such that the model is mathematically consistent over different data types and in terms of water balance.

In Paper A we explore if it is possible to increase the predictability of annual runoff by including point referenced precipitation and evaporation data in the geostatistical runoff modeling. In the model we propose, precipitation and evaporation data are used together with areal referenced streamflow observations. The areal referenced runoff observations are modeled as described in Section 2.1, while the precipitation and evaporation data are included by using the basic water balance equation in (1): The observed point runoff is the observed evaporation extracted from the observed precipitation, assuming storage effects are 0. The latter is realistic on an annual scale, for a hydrological year. Recall that the hydrological year is defined such that the precipitation that came as snow during the year has melted before the beginning of the next hydrological year, minimizing storage effects. Based on the above, we specify an observations

likelihood for point referenced annual runoff y_k as follows:

$$y_k = P_k - E_k = q(\boldsymbol{u}_k) + \epsilon_k^y \tag{22}$$
$$\pi(y_k|q(\boldsymbol{u}_k), \sigma_y) \sim \mathcal{N}(q(\boldsymbol{u}_k), s_k^y \sigma_y^2),$$

where $q(\boldsymbol{u}_k)$ is the underlying true runoff at point location $\boldsymbol{u}_k$ and the ϵ_k^y's are independent and identically distributed error terms with prior $\mathcal{N}(0, s_k^y \sigma_y^2)$ where $\{s_k^y\}$ are fixed predetermined scales and σ_y^2 is a variance parameter. The variables P_k and E_k are the observed precipitation and evaporation respectively. The true point runoff $q(\boldsymbol{u})$ is given by a stationary and continuous spatial model that is further specified in Section 2.3.

A combined point and areal model can next be achieved by multiplying the likelihood for the point observations (Equation (22)) with the likelihood for the areal observations (Equation (20)) for all target locations and catchments. Both observation likelihoods are given conditional on the same underlying stationary spatial model $q(\boldsymbol{u})$; for the areal model implicitly through $Q(\mathcal{A})$ which is given by a weighted sum of point runoff $q(\boldsymbol{u})$, according to Equation (19). This model formulation gives a mathematically consistent runoff model for which the water balance is preserved for the predicted runoff at any point in the study area.

Furthermore, to influence the model to weight the streamflow observations more than the more uncertain point observations, knowledge-based priors are used for the measurement uncertainties $s_i^z \sigma_z^2$ and $s_k^y \sigma_y^2$ in Equation (20) and Equation (22): In Paper A we specify a PC prior with a lower prior mean observation uncertainty for the areal runoff $s_i^z \sigma_z^2$, compared to the prior observation uncertainty for the point runoff $s_k^y \sigma_y^2$ (relative to the observed value). The prior credible interval for areal runoff $s_i^z \sigma_z^2$ is also more narrow. The scales s_k^y and s_i^z are included such that it is possible to let the prior (and posterior) measurement uncertainties increase with the observed value, which is a reasonable assumption for hydrological data. This way hydrological assumptions are incorporated into the model through knowledge-based prior distributions.

In the statistics literature, there exist other Bayesian geostatistical models that combines point and areal observations. Here, it is the model in

Moraga et al. (2017) that is most similar to the work presented here. Also these authors build a model for which the spatial variable at areal level is defined as a spatial average of the variable at point level, as in Equation (19). Other similarities are that their model is a Bayesian hierarchical model and that they use INLA and SPDE to make inference and predictions. However, they use a different model for the linear predictor $q(\boldsymbol{u})$ than we do, and they demonstrate their model by simulation studies and by modeling the spatial risk pattern for respiratory diseases and lung cancer in Zürich. Another Bayesian geostatistical model that combines point and areal referenced data is found in Wang et al. (2018). Here, the Bayesian model is fitted by using `Rstan` and demonstrated for epidemiological data. Furthermore, in Goovaerts (2010), a Kriging approach that combines point and areal data is presented. The papers by Moraga et al. (2017); Wang et al. (2018); Goovaerts (2010) all conclude that a combination of point and areal data can lead to a better predictive performance compared to when using one of the data types alone, but it requires that the observation types are compatible and modeled in a consistent way. Our results in Paper A support their conclusions: We find that when we predict runoff by only using point observations of runoff according to Equation (22), this gives poor results. Predicting runoff by only using areal observations gives better predictions. However, when combining the point observations with areal referenced streamflow observations, it leads to an increased predictive performance on average, compared to using one of the data types alone. Some of the increase in predictive performance can be explained by the mass conserving properties of the model, illustrated in Section 2.1.

In spite of the promising results in Paper A, precipitation and evaporation data were not included in the runoff interpolation in Paper B. For Paper B we had substantially more streamflow data available, and the precipitation and evaporation data turned out to contribute to a (small) decrease in predictive performance compared to when not using them. Consequently we decided to omit the precipitation and evaporation data from the analysis in Paper B. Our experience is that precipitation and evaporation data can improve the runoff predictions in areas with few streamflow observations. However, when there are more streamflow data available, the precipitation data are redundant and/or can affect the predictions negatively in our proposed model. Based on this, more research should be done around this

topic, to find the best way to connect precipitation and evaporation data to runoff in a geostatistical model. A key challenge here is probably to find a clever way to account for that precipitation gauges often are located a relatively low elevations. Hence, there is often a lack of information about the precipitation field at high elevations. At the same time, we typically find the most extreme precipitation and the largest spatial variability in mountainous areas.

2.3 Exploiting short records - Introducing the two field model (Paper A-B)

Hydrologists are often interested in the long-term behavior of a catchment, and mean annual runoff is a key hydrological quantity (Blöschl et al., 2013). The mean annual runoff is typically calculated based on a 30 year period, as an average over 30 years represents a climate (WMO, 1992). Calculating this is straight forward for catchments where there are 30 years of observations available. However, most catchments in the world don't have runoff observations or only have so-called short records of annual runoff which means that the catchments only have a few years of data available. We refer to these catchments as *partially gauged*.

As mentioned in the introduction, short records have been omitted from analyses of mean annual runoff in Norway (Beldring et al., 2002). The reason is that averages based on 1-29 years of data are regarded as unreliable estimates of the true mean annual runoff for the 30 year target period. However, omitting partially gauged catchments from an analysis means that a lot of potential information might be lost. Hence, research on how much information that actually is stored in the Norwegian short records and how to include them in the runoff modeling, is highly relevant.

In the literature there exist several approaches for exploiting short records of runoff data. These are known as record augmentation techniques. Assume that a partially gauged catchment has n annual runoff observations $y_1, .., y_n$ from n years. The first step in most record augmentation methods is to find one or several so-called donor catchments for the target catchment. These should have longer time series of runoff than the target catchment, and for a time period that overlaps with the observations

of $y_1, ..., y_n$. The donor catchments are selected based on proximity in space and/or catchment similarity, i.e. that the donor and target catchments have similar vegetation, precipitation patterns, slope, elevation or other catchment characteristics. The idea is that catchments with similar characteristics also have similar runoff patterns.

After selecting suitable donors, the next step is to develop a relationship between the target catchment and the donor catchment(s). Here, several approaches are available. An obvious approach is to develop a linear relationship $y_i = \beta_0 + \beta_1 x_i$ by simple linear regression, where y_i and x_i are observations from the target and donor catchment respectively, and β_0 and β_1 are regression coefficients that must be estimated. A slightly more sophisticated approach is to use MOVE methods. Here, the coefficients β_0 and β_1 are estimated by assuming that the sample mean and the sample variance of the target catchment are maintained over time. The MOVE estimators are shown to have better properties than the linear regression estimators (Hirsch, 1982).

In addition to the above methods, there exist a variety of other record augmentation procedures. Many of them are based on utilizing the correlation between the target catchment and the donor catchment(s) and/or on finding a suitable ratio for weighting the time series of runoff between the target and the donor relative to each other. We refer to some of the approaches here: Fiering (1963); Laaha and Blöschl (2005); Vogel and Stedinger (1985); Matalas and Jacobs (1964).

In Paper A and Paper B we present spatial models for annual runoff that also can be considered as record augmentation techniques. The spatial models are designed to be particularly suitable for datasets with many missing values. This is achieved by constructing a *two field model* that simultaneously models several years of runoff, and that includes a spatial component that is able to capture repeated runoff patterns over time. More specifically, assume that $q_j(\boldsymbol{u})$ is the true runoff in year j at a point location $\boldsymbol{u} \in \mathcal{D}$. The annual runoff is modeled as

$$q_j(\boldsymbol{u}) = \beta_c + c(\boldsymbol{u}) + \beta_j + x_j(\boldsymbol{u}); \quad j = 1, .., r, \tag{23}$$

conditional on the model parameters, where β_c is an intercept common for all years and $c(\boldsymbol{u})$ is a spatial field common for all years. These com-

ponents model long-term averages of runoff, or the climate in the study area, and we often refer to $c(\boldsymbol{u})$ as *the climatic spatial field*. Similarly are β_j and $x_j(\boldsymbol{u})$ a year specific intercept and a year specific spatial field respectively. These model the the annual runoff that cannot be explained by long-term patterns, i.e. the spatial variability due to more year dependent, random effects. The field $x_j(\boldsymbol{u})$ is referred to as *the year specific spatial field* or as *the replicated spatial field*. Both spatial fields are modeled as stationary Matérn GRFs given the underlying parameters: $c(\boldsymbol{u})$ with range parameter ρ_c and marginal variance σ_c^2, and $x_j(\boldsymbol{u})$ with range parameter ρ_x and marginal variance σ_x^2. The spatial fields $x_j(\boldsymbol{u})$ for $j = 1, ..r$ are regarded as independent realizations or replicates of the underlying GRF. The same applies for the year specific intercepts β_j for $j = 1, ..., r$ that are regarded as replicates of the same normal prior distribution. Consequently, Cov($x_j(\boldsymbol{u})$,$x_v(\boldsymbol{u})$)=0 and Cov(β_j,β_v)=0 conditional on the model parameters, as long as $j \neq v$.

The annual runoff in Equation (23) is observed through point and/or areal referenced observations with uncertainty, and the observation models are already discussed in Section 2.1 and 2.2. To achieve a joint runoff model for year $j = 1, ..r$, the observation likelihoods for the observed runoff are simply multiplied together for all years $j = 1, ...r$ and target locations and/or catchments.

It is the climatic part of the model in Equation (23) that makes the model particularly suitable for record augmentation. If there is a few annual observations available from the target catchment, the climatic part of the model captures the magnitude of runoff here compared to nearby locations. This information is transferred across time. Hence, if the study area is dominated by repeated runoff patterns explained by $c(\boldsymbol{u})$ with $\sigma_c \gg \sigma_x$, short records can have a large impact on the predictions for all study years. On the other hand, if $x_j(\boldsymbol{u})$ is the dominating spatial effect with $\sigma_c \ll \sigma_x$, the proposed model behaves as a standard spatial model: The short records will only have an impact on the results for the years where we have data.

To demonstrate when the suggested model is particularly useful, we show an example in Figure 3. Here, we see time series of annual runoff from 8 Norwegian catchments from 1996-2005. What is interesting in this dataset, is that the spatial variability is high. The variability over time is also quite

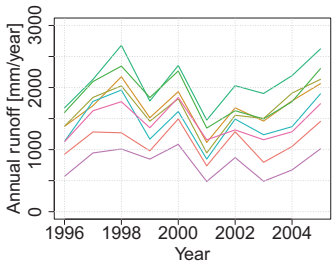


Figure 3: Time series of annual runoff from 8 Norwegian catchments from 1996-2005.

large. However, the time series of annual runoff are almost parallel over time and the ranking between stations (from wet to dry) is approximately constant. The latter indicates that the annual runoff follows a strong, repeated spatial pattern. Consequently, if we have information from a catchment for one year, this tells a lot about the runoff at this location relative to neighboring locations for other years as well, in spite of the large spatial and temporal variability.

The Norwegian annual runoff tends to follow patterns like we see in Figure 3, and in Paper B we show that including a short record of length one for a target catchment, can lead to a 50 % reduction in RMSE in the proposed model compared to a situation where we have no data from the target catchment. This result tells us two things: 1) The information NVE looses by not including short records in the modeling of annual runoff in Norway is substantial and 2) the proposed model is able to capture and exploit the kind of pattern we see in Figure 3.

The benefit of the model in Equation (23) as a record augmentation procedure, is that it is a flexible model that borrows strength both in space and time. It is able to detect long-term spatial variability and adjusts the two spatial fields $c(\boldsymbol{u})$ and $x_j(\boldsymbol{u})$ relative to each other. Consequently, it is quite risk-free to include (very) short records in the modeling. This is different from other hydrological models where it has been looked upon as risky to include short records of length e.g. 1-5 years in the modeling of long-term averages of 30 years. Another property of the proposed model

is that it allows short records of length one. Other record augmentation procedures like linear regression and correlation based methods require at least two pairs of observations from the target catchment and the donor catchment to give meaningful results. In the proposed model, we also avoid the challenge of selecting suitable donor catchments, which is a key challenge for other comparable record augmentation methods.

Comparing the proposed model to geostatistical models used for runoff interpolation, these are typically not able to exploit short records. In Top-Kriging it is possible to include short records in the interpolation of mean annual runoff by simply using the average of the short record as an approximation for the long-term mean annual runoff. Next, this observation's uncertainty is set based on record length (Skøien et al., 2006). However, this approach should be more sensitive to the record length and to extreme observations than our approach. Furthermore, it is not straight forward how short records can be incorporated into Top-Kriging if we are interested in predicting runoff for individual years. Hence, the main novelty of the model in Equation (23) is its property to exploit short records within a geostationary framework in a robust and relatively risk-free way.

A model class in statistics that is based on similar ideas as the two field model in Equation (23), is co-regionalization models (see e.g. Schmidt and Gelfand (2003); Wackernagel (1998); Gelfand et al. (2010)). A co-regionalization model can be used to build a joint model for two or more variables that don't need to be measured at the same spatial locations, e.g. precipitation, runoff and elevation (Adhikary et al., 2017). The target variables share a spatial component that quantifies the cross-correlation between them. When the cross-correlation is large, the co-regionalization model can improve the predictability of the target variables compared to when modeling each of the variables separately. In the model in Equation (23), we can we think of the spatial component $c(\boldsymbol{u})$ as a component that exploits cross-correlation. However, this cross-correlation is computed between different locations *across time*, rather than across different target variables. Hence, we only need one data type for using the two field model (as in Paper B where we model areal runoff), but we have also shown that it is possible to combine two data types (as in Paper A where we model point runoff and areal runoff together). Our model however, re-

quires spatio-temporal data to be able to exploit cross-correlation, unlike co-regionalization models that can be built both for spatial and spatio-temporal data.

Since the two field model is made for analyzing spatio-temporal data it also makes sense to compare it to spatio-temporal models. Compared to a standard separable spatio-temporal model that changes in time according to a first-order autoregressive process (Blangiardo and Cameletti, 2015), the suggested two field model is considerably faster to run. The two field model also gives a different uncertainty model that might be more realistic when studying long-term averages and variables that follow repeated spatial patterns. More discussion around this topic can be found in the Technical note.

2.4 Incorporating process-based information into a geostatistical framework (Paper C)

In additional to statistical approaches, there is one other main class of approaches to runoff estimation, namely process-based approaches. In process-based models, fundamentals laws of physics are used to estimate runoff. These models are typically built on equations that connect runoff to variables like precipitation, temperature, vegetation and land use, and account for processes like snow storage, snow melt, evapotranspiration and soil infiltration. Conservation laws like preservation of mass and energy are used to e.g. ensure that the water balance is preserved over nested catchments. Furthermore, runoff data from gauged catchments are typically used to calibrate model parameters such that the global error between the model and the observed runoff is minimized (Beldring et al., 2003; Doherty, 2004; Lawrence et al., 2009). Some process-based models that are commonly used in Norway are the HBV model (Sveriges Meteorologiska och Hydrologiska Institut, 1992) and WASMOD (Widén-Nilsson et al., 2007).

The strength of process-based models is their ability to take mass balance properties into account and connect the variables together in a consistent way. Geostatistical models are on the other hand typically better at providing a good fit between the model and the data in areas where there are

observations available, and also provide uncertainty quantification. Motivated by the strength and weaknesses of these two model classes, there exist several attempts on combining process-based models and geostatistics in the hydrology literature. For example, in Pannecoucke et al. (2020) a process-based model was used to simulate flow a number of times. Next, the results were used to build an empirical variogram to be used for Kriging. The approach was successfully used to estimate the level of contamination within the soil. In Rivest et al. (2008) a simple conceptual flow model was solved numerically, and the results were next used for external drift Kriging. External drift Kriging was also used in Laaha et al. (2013) to interpolate streamflow temperature. Here, a physical relationship between mean annual stream temperature and stream gauge altitude was incorporated into the model and combined with the Top-Kriging approach.

The idea behind the methods mentioned above, is that a combination model performs better than a geostatistical or process-based model alone. This was also confirmed by their results. However, few combination models for mean annual runoff were found in our literature search. Some models exist, for example the model in Qiu et al. (2018) where the authors combined a process-based Budyko water balance model with a geostatistical approach. In Sauquet (2006) mean annual runoff was estimated by a Kriging approach that is able to incorporate basin characteristics through a function $g(\cdot)$ and residual Kriging. The method was demonstrated with catchment elevation as input to $g(\cdot)$, but more complex covariates can be included. Operationally in Norway, mean annual runoff has been estimated by using the process-based HBV model. Next, the ratio between the observed runoff and the runoff estimated by the HBV model has been interpolated from gauged to ungauged catchments by using an inverse distance method Beldring et al. (2002). This is a non-statistical method without uncertainty quantification (see e.g. Dingman (2015); Lu and Wong (2008)). However the inverse distance interpolation could easily be replaced by geostatistical interpolation, resulting in a combined process-based and geostatistical approach for mean annual runoff.

In Paper C we present a model for mean annual runoff that incorporates the simulations from a process-based model into a geostatistical framework. This is done by using mean annual runoff as a response variable and

simulations of runoff from a process-based model as a covariate. The covariate is available on a regular grid for the area of interest. In our model, we assume that the ratio between the process-based runoff simulations and the true streamflow varies in the study area. This is achieved by connecting the response variable and the covariate through a regression coefficient that is allowed to vary in space, hence it is a spatially varying coefficient (SVC).

A variety of applications of spatially varying coefficients are reported in the literature (Gelfand et al., 2003; Ferguson et al., 2009; Hastie and Tibshirani, 1993; Finley, 2011; Lu et al., 2009; Su et al., 2017). There are also several ways a SVC model can be implemented. The simplest option is to divide the study area into regions and let each region have its own regression coefficient like in e.g. Gamerman et al. (2003). However, how the study area should be partitioned is not always intuitive. Another approach is found in Gelfand et al. (2003). Here, a SVC is introduced by modeling the regression coefficient as a Gaussian random field. Hence, the relationship between the response and the covariate follows a dependency structure defined by a GRF and its covariance function.

In Paper C we adopt the approach from Gelfand et al. (2003) and use it to interpolate the ratio between the process-based runoff estimate and the observed streamflow, from gauged catchments to ungauged catchments. The model also includes an additional spatial field. More specifically the model is given by

$$q(\boldsymbol{u}) = \beta_0 + (\beta_1 + \alpha(\boldsymbol{u})) \cdot h(\boldsymbol{u}) + x(\boldsymbol{u}), \tag{24}$$

where $q(\boldsymbol{u})$ is the true mean annual runoff at a location or grid cell $\boldsymbol{u} \in \mathcal{R}^2$, $h(\boldsymbol{u})$ is a simulation from a process-based model for the same grid cell $\boldsymbol{u}$, and $\alpha(\boldsymbol{u})$ is a GRF that together with the fixed effect β_1 defines the spatially varying coefficient. Further is the variable β_0 an intercept and $x(\boldsymbol{u})$ is another GRF. The model is given conditional on model parameters that are given informative prior distributions. The true streamflow in a catchment $\mathcal{A}_i$ is observed through noisy areal referenced observations z_i, similarly as in Equation (19) and Equation (20). Hence, the nested structure of catchments is taken into account, and we interpolate data between catchments areas and point referenced grid cells in a consistent way.

In Beldring et al. (2002) the ratio $z_i/h(\mathcal{A}_i)$ between the simulated streamflow $h(\mathcal{A}_i)$ from a process-based model, and the observed streamflow from gauged catchments z_i, were interpolated to improve the mean annual runoff estimates in Norway. The model from Equation (24) hence does something similar through the spatially varying coefficient $\beta_1 + \alpha(\boldsymbol{u})$ as it models the spatial structure of the ratio between the streamflow and the process-based simulations in the study area. However, as we use a geostatistical approach we also obtain uncertainty quantification, differently from Beldring et al. (2002) where they used a non-statistical inverse-distance approach. Furthermore, the model in Equation (24) is able to account for one additional dependency structure between the process-based estimate and the streamflow, through the GRF $x(\boldsymbol{u})$. The dependency structure captured by $x(\boldsymbol{u})$ is not so different from what we would obtain from performing residual interpolation. i.e. interpolation of $z_i - h(\mathcal{A}_i)$. Residual interpolation is another approach used in hydrology to improve the results from an initial model (see e.g. the *Georegression* approach in Merz and Blöschl (2005)).

In Paper C, the so-called SVC model from Equation (24) was used to predict mean annual runoff in ungauged and partially gauged catchments in Norway, and simulations from the HBV model were used as the covariate $h(\cdot)$. The two field model from Equation (23) was used as a preprocessing step for record augmentation of short records before further analysis with the SVC model: This way we were able to include information from both fully gauged catchments and partially gauged catchments in the observation likelihood. Our findings in Paper C support the conclusions from the work by Pannecoucke et al. (2020); Rivest et al. (2008); Qiu et al. (2018): The combination model from Equation (24) performed better than a purely geostatistical approach (Top-Kriging and the two field model) and a purely process-based model (the HBV model) when making predictions for ungauged catchments. For partially gauged catchments however, the two field model from Equation (23) gave better predictions than the SVC model. It is not surprising that a purely data-driven method performs better than a method with explanatory variables at locations where there are a few years of data available. This is one of the strengths of purely geostatistical methods. The results in Paper C also confirmed that the two field model can be used as a preprocessing step for partially gauged catchments before further analysis with other models (such as the SVC model).

3 Overview of the scientific work and suggested areas of use beyond annual runoff interpolation

Section 2 explains the contributions of the thesis towards annual runoff estimation. We now discuss contributions outside the field of annual runoff interpolation and suggest topics for further research. We start by giving a brief overview of the contents of Paper A-C and the Technical note.

Paper A: We develop a two field model for annual runoff that supports both areal referenced streamflow observations of runoff and point referenced precipitation and evaporation observations. The model is tested on a small dataset from western Norway, and we demonstrate the model's mass preserving properties by a real case example. Through a simulation study we illustrate how the parameters of the two field model can be used to understand systematic prediction bias over time. The paper is written for statisticians.

Paper B: We demonstrate how the two field model can be used for interpolation of (mean) annual runoff, and the main focus is on showing how the model is particularly useful for exploiting short records. This time we only use streamflow observations of runoff (not precipitation and evaporation data), and compare the results to Top-Kriging. The model is evaluated for a larger dataset consisting of catchments from all over Norway. The paper is written for hydrologists.

Paper C: In this paper we incorporate process-based hydrological simulations into a geostatistical framework for mean annual runoff. This is done by developing a linear relationship between the observed mean annual runoff (response variable) and a gridded runoff product that is simulated by a process-based hydrological model (covariate). The linear relationship is allowed to vary in space according to a GRF, i.e. the regression coefficient is a spatially varying coefficient (SVC). The two field model from Paper A and B is used as a preprocessing step for record augmentation for the partially gauged catchments before further analysis with the SVC model. The paper is written for hydrologists.

Technical note: We investigate properties of the two field model that were considered outside the scope of Paper A and B. In the note, the

two field model is compared to a simpler one field spatial model and to a comparable spatio-temporal model. We illustrate when and why the two field model can be useful and provide a collection of case examples. Through simulation studies, general properties of the model are explored, such as parameter identifiability. The note also contributes with `r-inla` code for implementing the two field model. It is written for statisticians.

An important part of the field of statistics is to contribute with methods for data analysis for other scientific disciplines, and in this thesis we combine different state-of-the-art statistical methods to develop new approaches for runoff interpolation. The models introduced in Paper A-C all have the same methodological foundation: They are three-staged geostatistical Bayesian hierarchical models that use two GRFs to model the process of interest. Newly developed priors (PC priors) are used to incorporate expert knowledge. To increase the predictability of runoff, several data sources are combined, e.g preprocessed short records, point and areal referenced data and simulations from a process-based model, giving a variety of approaches for spatial interpolation. Due to developments like INLA and the SPDE approach to spatial modeling, we are able to not only run the proposed models, but also to perform large cross-validation assessments on them.

In the above paper summaries, we see that all papers utilize the two field model from Equation (23), and that each work presents a new aspect of it. Hence, in addition to contributing to improving the existing runoff models, the thesis offers a comprehensive study of the two field model which turns out to have several interesting properties. The Technical note was included in the thesis to illustrate the generality of the work: It is not only runoff that follows repeated spatial trends over time. The patterns we see in Figure 3 also occur for other environmental variables, typically due to repeated wind patterns and the topography in the study area. This is the case particularly for environmental variables measured over longer temporal scales, such as annual and monthly data. Hence, the two field model could be useful outside the field of runoff interpolation, and we present a collection of case examples in the Technical note.

In the Technical note we also illustrate what happens if we use a simple spatial model or a spatio-temporal model for the kind of data we see in Figure 3. This was done to demonstrate when the two field model has benefits

over comparable models. In short, the two field model has useful properties when modeling environmental variables that are driven by repeated spatial patterns over time, and the available dataset has several missing values. It is also particularly suitable for modeling long-term averages. In a setting like this a spatial model is not able to transfer information from short records across time, while a spatio-temporal model typically assumes an increasing uncertainty over time. The latter is not necessarily realistic when modeling long-term averages for variables that follow strong, repeated spatial patterns like the data in Figure 3.

In Paper A we present a simulation study on how the parameters of the climatic spatial field σ_c, ρ_c and the parameters of the replicated spatial field σ_x, ρ_x from Equation (23) can provide interesting information about the study area and the variable of interest. To give an example; assume that it is of interest to interpolate annual precipitation for a 10 year period to a location without observations. The predictions are done based on observations from nearby precipitation gauges. If the spatial pattern of precipitation mainly comes from long-term patterns (such as in Figure 3) and the spatial range ρ_c is low, it is a relatively large probability that the precipitation predictions for the target location are systematically biased over time: Strong long-term patterns means that the weather pattern in the area repeats itself ($\sigma_c \gg \sigma_x$), implying that also prediction errors tend to repeat themselves, as long as the observation set-up is unchanged. Hence, a poor precipitation prediction in year j probably means that you obtain a poor precipitation prediction for all other years as well, unless new measurement locations become available. This example illustrates that two field model and its parameters can provide an understanding about the predictive biases and uncertainties we see in a study area. It can for example be used as a tool for understanding whether prediction biases are due to repeated long-term patterns of the underlying variable, or due to mismatch between the model and the data. See Paper A for more.

Furthermore, in Paper B we show how (very) short records often have a large impact on the predictions for a target location in areas where most of the spatial variability can be explained by long-term effects, i.e. in areas where $\sigma_c \gg \sigma_x$. Again, particularly if the climatic range ρ_c is small

relative to the spatial variability. This was the case for the Norwegian annual runoff, and also for other environmental variables that we have tested. Motivated by this finding, the parameters of the two field model could be used by decision-makers to indicate the value of installing a new measuring station for an environmental variable of interest. A new measuring station could improve the long-term estimates of the target variable only a year after installation for certain climates, as demonstrated in Paper B. For a study area with such properties, the cost of setting up a new measuring station might be lower than the potential cost of a poor prediction. Similarly could the two field model also be used to evaluate whether a measuring station could be shut down to save money, motivated by that we already know enough about the spatial variability for that location (relative to neighboring stations). The above examples are related to the field of decision theory and assessing the value of information (Eidsvik et al., 2015). Utilizing the two field model and its parameters to find optimal observation designs might be an interesting topic for further research.

One of the most essential tasks for hydrologists is to build models for flood and flood forecasting. This is important for building robust infrastructure and for our security. Hydrologists typically model flood and their return periods by using generalized extreme value distributions (GEV), such as the Gumbel distribution. In Norwegian flood models, explanatory variables are used to estimate the location parameter in the selected GEV distribution (Engeland et al., 2020; Thorarinsdottir et al., 2018), and in the currently used NVE models described in Engeland et al. (2020), mean annual runoff is the explanatory variable that is weighted the most. In this thesis, we have demonstrated how a few annual runoff observations can contribute to substantial improvements in the long-term estimates of mean annual runoff, using the proposed two field model. Based on this, a logical next step would be to investigate if a few annual observations also can improve the accuracy of flood models, implicitly through improved mean annual runoff estimates. This is another relevant topic for further research.

References

K. Adamowski and C. Bocci. Geostatistical regional trend detection in river flow data. *Hydrological Processes*, 15(18):3331–3341, 2001. doi: 10.1002/hyp.1045.

S. K. Adhikary, N. Muttil, and A. G. Yilmaz. Cokriging for enhanced spatial interpolation of rainfall in two australian catchments. *Hydrological Processes*, 31(12): 2143–2161, 2017. doi: 10.1002/hyp.11163. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.11163`.

H. Bakka, H. Rue, G-A. Fuglstad, A. Riebler, D. Bolin, J. Illian, E. Krainski, D. Simpson, and F. Lindgren. Spatial modeling with r-inla: A review. *WIREs Computational Statistics*, 10(6), 2018. doi: 10.1002/wics.1443.

H. Bakka, J. Vanhatalo, J. B. Illian, D. Simpson, and H. Rue. Non-stationary gaussian models with physical barriers. *Spatial Statistics*, 29:268 – 288, 2019. ISSN 2211-6753. doi: https://doi.org/10.1016/j.spasta.2019.01.002.

S. Banerjee, A.E. Gelfand, and B.P Carlin. *Hierarchical Modeling and Analysis for Spatial Data*, volume 101 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 2004.

S. Banerjee, A. E. Gelfand, A. O. Finley, and H. Sang. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(4):825–848, 2008. doi: 10.1111/j.1467-9868.2008.00663.x.

S. Beldring, L. A Roald, and A. Voksø. Arenningskart for Norge. *Tech. Rep. Oslo: NVE*, 2002. ISSN 1501-2840.

S. Beldring, K. Engeland, L. A. Roald, N. R. Sælthun, and A. Voksø. Estimation of parameters in a distributed precipitation-runoff model for norway. *Hydrology and Earth System Sciences*, 7(3):304–316, 2003.

E. Bellier, P. Monestiez, and C. Guinet. *Geostatistical Modelling of Wildlife Populations: A Non-stationary Hierarchical Model for Count Data*, pages 1–12. 07 2010. doi: 10.1007/978-90-481-2322-3_1.

J.O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York, 1985.

O. Berke. Exploratory disease mapping: Kriging the spatial risk function from regional count data. *International journal of health geographics*, 3:18, 09 2004. doi: 10.1186/1476-072X-3-18.

J. Bernardo. Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society. Series B*, 41, 01 1979. doi: 10.1111/j.2517-6161.1979.tb01066.x.

M. Blangiardo and M. Cameletti. *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Wiley, 1st edition, 2015.

G. Blöschl, M. Sivapalan, T. Wagener, A. Viglione, and H. Savenije. *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*. Camebridge University press, 2013.

S. Brenner and L. Scott. *The Mathematical Theory of Finite Element Methods, 3rd Edition. Vol. 15 of Texts in Applied Mathematics*. Springer, 2008.

B. Carpenter, A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. Stan : A probabilistic programming language. *Journal of Statistical Software*, 76, 01 2017. doi: 10.18637/jss.v076.i01.

G. Casella and R.L. Berger. *Statistical Inference*. Duxbury Press Belmont, 1990.

G. Casella and E. George. Explaning the gibbs sampler. *The American Statistician*, 46: 167–174, 08 1992. doi: 10.1080/00031305.1992.10475878.

A. Clements, N. Lwambo, L. Blair, U. Nyandindi, G. Kaatano, S. Kinung'hi, J. Webster, A. Fenwick, and S. Brooker. Bayesian spatial analysis and disease mapping: Tools to enhance planning and implementation of a schistosomiasis control programme in tanzania. *Tropical medicine & international health : TM & IH*, 11:490–503, 05 2006. doi: 10.1111/j.1365-3156.2006.01594.x.

N. Cressie. *Statistics for spatial data*. J. Wiley & Sons, 1993.

N. Cressie and G. Johannesson. Fixed rank kriging for very large data sets. *Journal of the Royal Statistical Society Series B*, 70:209–226, 02 2008. doi: 10.1111/j.1467-9868.2007.00633.x.

P. Diggle and P. Ribeiro Jr. *Model-Based Geostatistics*, volume 846. 01 2007. doi: 10.1007/978-0-387-48536-2.

S.L. Dingman. *Physical Hydrology*. Waveland Press, Incorporated, 2015. ISBN 9781478611189.

J. Doherty. Pest: Model independent parameter estimation. fifth edition of user manual. *Watermark Numerical Computing, Brisbane, Australia*, 2004.

J. Eidsvik, A. O. Finley, S. Banerjee, and H. Rue. Approximate bayesian inference for large spatial datasets using predictive process models. *Computational Statistics & Data Analysis*, 56(6):1362 – 1380, 2012. doi: 10.1016/j.csda.2011.10.022.

J. Eidsvik, T. Mukerji, and D. Bhattacharjya. *Value of Information in the Earth Sciences: Integrating Spatial Modeling and Decision Analysis*. Cambridge University Press, 2015. doi: 10.1017/CBO9781139628785.

K. Engeland, P. Glad, H. B. Hamududu, H. Li, T. Reitan, and S. M. Stenius. Lokal og regional flomfrekvensanalyse. *Tech. Rep. Oslo: NVE*, 2020.

C. A. Ferguson, A. W. Bowman, E. M. Scott, and L. Carvalho. Multivariate varying-coefficient models for an ecological system. *Environmetrics*, 20(4):460–476, 2009. doi: 10.1002/env.945.

E. Ferkingstad and H. Rue. Improving the inla approach for approximate bayesian inference for latent gaussian models. *Electronic Journal of Statistics*, 9, 12 2015. doi: 10.1214/15-EJS1092.

M.B. Fiering. Use of correlation to improve estimates of the mean and variance. *USGS Publications Warehouse*, 1963.

A. O. Finley. Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, 2(2):143–154, 2011. doi: 10.1111/j.2041-210X.2010.00060.x.

Y. Fong, H. Rue, and J. Wakefield. Bayesian inference for generalized linear mixed models. *Biostatistics (Oxford, England)*, 11:397–412, 12 2009. doi: 10.1093/biostatistics/kxp053.

G-A. Fuglstad, I. Hem, A. Knight, H. Rue, and A. Riebler. Intuitive joint priors for variance parameters. *Bayesian Analysis*, 10 2019a. doi: 10.1214/19-BA1185.

G-A. Fuglstad, D. Simpson, F. Lindgren, and H. Rue. Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association*, 114(525):445–452, 2019b. doi: 10.1080/01621459.2017.1415907.

R. Furrer, M. Genton, and D. Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15, 12 2005. doi: 10.1198/106186006X132178.

D. Gamerman and H. F Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC, 2006.

D. Gamerman, A.R.B. Moreira, and H. Rue. Space-varying regression models: specifications and simulation. *Computational Statistics and Data Analysis*, 42(3):513 – 533, 2003. ISSN 0167-9473. doi: https://doi.org/10.1016/S0167-9473(02)00211-6. Computational Econometrics.

L.S. Gandin. On optimal interpolation and extrapolation of meterological fields. *Trudy Main Geophys. Obs.*, 114:75–89, 1960.

A. Gelfand, P. Diggle, M. Fuentes, and P. Guttorp. Handbook of spatial statistics. *Chapman & Hall*, 01 2010. doi: 10.1201/9781420072884.

A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

A.E. Gelfand, K. Hyon-Jung, C.F. Sirmans, and S. Banerjee. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98:387–396, 02 2003. doi: 10.2307/30045248.

A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1, 2006. doi: 10.1214/06-BA117A.

A. Gelman and J. Hill. *Data analysis using regression and multilevel/hierarchical models*, volume Analytical methods for social research. Cambridge University Press, New York, 2007.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd ed. edition, 2004.

P. Goovaerts. *Geostatistics for Natural Resource Evaluation*, volume 42. 01 1997.

P. Goovaerts. Combining areal and point data in geostatistical interpolation: Applications to soil science and medical geography. *Mathematical geosciences*, 42:535–554, 07 2010. doi: 10.1007/s11004-010-9286-5.

L. Gottschalk. Correlation and covariance of runoff. *Stochastic Hydrology and Hydraulics*, 7:85–101, 1993a. doi: 10.1007/BF01581418.

L. Gottschalk. Interpolation of runoff applying objective methods. *Stochastic Hydrology and Hydraulics*, 7:269–281, 1993b.

G. Guillot, R. Vitalis, A. le Rouzic, and M. Gautier. Detecting correlation between allele frequencies and environmental variables as a signature of selection. a fast computational approach for genome-wide studies. *Spatial Statistics*, 8:145 – 155, 2014. ISSN 2211-6753. doi: 10.1016/j.spasta.2013.08.001.

P. Guttorp and T. Gneiting. Studies in the history of probability and statistics XLIX on the matérn correlation family. *Biometrika*, 93(4):989–995, 2006. doi: 10.1093/biomet/93.4.989.

T. Hastie and R. Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(4):757–779, 1993. doi: 10.1111/j.2517-6161.1993.tb01939.x.

R. M. Hirsch. A comparison of four record extension techniques. *Water Resour. Res.*, 1982.

J. Huang, B. Malone, B. Minasny, A. Mcbratney, and J. Triantafilis. Evaluating a bayesian modelling approach (inla-spde) for environmental mapping. *Science of The Total Environment*, 609:621–632, 12 2017. doi: 10.1016/j.scitotenv.2017.07.201.

R. Ingebrigtsen, F. Lindgren, and I. Steinsland. Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, 8:20 – 38, 2014. doi: 10.1016/j.spasta.2013.06.002.

R. Ingebrigtsen, F. Lindgren, I. Steinsland, and S. Martino. Estimation of a non-stationary model for annual precipitation in southern Norway using replicates of the spatial field. *Spatial Statistics*, 14:338 – 364, 2015. doi: 10.1016/j.spasta.2015.07.003.

H. Jeffreys. An invariant form for the prior probability in estimation problems. 186 (1007):453–461, 1946. ISSN 00804630. doi: 10.2307/97883.

M. Jullum, T. Thorarinsdottir, and F. E. Bachl. Estimating seal pup production in the greenland sea by using bayesian hierarchical modelling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(2):327–352, 2020. doi: 10.1111/rssc.12397. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssc.12397`.

R. Kerry, P. Goovaerts, R. Haining, and V. Ceccato. Applying geostatistical analysis to crime data: Car-related thefts in the baltic states. *Geographical analysis*, 42:53–77, 01 2010. doi: 10.1111/j.1538-4632.2010.00782.x.

D. Khan and M. Warner. A bayesian spatial and temporal modeling approach to mapping geographic variation in mortality rates for subnational areas with r-inla. *Journal of data science: JDS*, 18:147–182, 01 2018.

E. Krainski, V. Gómez-Rubio, H. Bakka, A. Lenzi, D. Castro-Camilo, D. Simpson, F. Lindgren, and H. Rue. *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. 09 2018. doi: 10.1201/9780429031892.

D. G. Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 56(6):119–139, 1966.

P. Kyriakidis, J. Kim, and N. Miller. Geostatistical mapping of precipitation from rain gauge data using atmospheric and terrain characteristics. *Journal of Applied Meteorology - J APPL METEOROL*, 40:1855–1877, 11 2001. doi: 10.1175/1520-0450(2001)040<1855:GMOPFR>2.0.CO;2.

G. Laaha and G. Blöschl. Low flow estimates from short stream flow records — a comparison of methods. *Journal of Hydrology*, 306(1):264 – 286, 2005. doi: 10.1016/j.jhydrol.2004.09.012.

G. Laaha, J. Skøien, F. Nobilis, and G. Blöschl. Spatial prediction of stream temperatures using top-kriging with an external drift. *Environmental Modeling & Assessment*, 18, 12 2013. doi: 10.1007/s10666-013-9373-3.

D. Lawrence, I. Haddeland, and E. Langsholt. Calibration of hbv hydrological models using pest parameter estimation. *Tech. Rep. Oslo: NVE*, 2009.

A. Lenzi, I. Steinsland, and P. Pinson. Benefits of spatiotemporal modeling for short-term wind power forecasting at both individual and aggregated levels. *Environmetrics*, 29(3):e2493, 2018. doi: 10.1002/env.2493. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2493`. e2493 env.2493.

E. Lesaffre and A. Lawson. *Bayesian Biostatistics*, volume 25. 07 2012. doi: 10.1016/S0169-7161(05)25025-3.

M.N.M. Lieshout. *Theory of Spatial Statistics: A Concise Introduction*. 03 2019. ISBN 9780429052866. doi: 10.1201/9780429052866.

F. Lindgren, H. Rue, and J. Lindström. 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)*, 73:423–498, 2011. doi: 10.1111/j.1467-9868.2011.00777.x.

G. Y. Lu and D. W. Wong. An adaptive inverse-distance weighting spatial interpolation technique. *Computers & Geosciences*, 34(9):1044 – 1055, 2008. ISSN 0098-3004. doi: https://doi.org/10.1016/j.cageo.2007.07.010.

Z. Lu, D. J Steinskog, D. Tjøstheim, and Q. Yao. Adaptively varying-coefficient spatiotemporal models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(4):859–880, 2009. doi: 10.1111/j.1467-9868.2009.00710.x.

S. Martino, R. Akerkar, and H. Rue. Approximate bayesian inference for survival models. *Scandinavian Journal of Statistics*, 38:514 – 528, 09 2011. doi: 10.1111/j.1467-9469.2010.00715.x.

N. C. Matalas and B. Jacobs. A correlation procedure for augmenting hydrologic data. *U.S. Geol. Surv. Prof. Pap.*, 434-E(E1-E7), 1964.

G. Mathéron. Applications des méthodes statistiques à l'evaluation des gisements. *Annales des Mines*, 12:50 – 75, 1955.

G. Mathéron. Principles of geostatistic. *Economic Geology*, 58:1246 – 1266, 1963.

B. Matérn. Spatial variation: Stochastic models and their application to some problems in forest surveys and other sampling investigations. *Meddelanden från Statens Skogforskningsinstitutt*, 49(5), 1960.

R. Merz and G. Blöschl. Flood frequency regionalisation - spatial proximity vs. catchment attributes. *Journal of Hydrology*, 302:283 – 306, 2005. doi: 10.1016/j.jhydrol.2004.07.018.

P. Moraga. *Geospatial Health Data: Modeling and Visualization with R-INLA and Shiny*. Chapman and Hall/CRC, 11 2019. ISBN 9780367357955.

P. Moraga, S. M. Cramb, K. L. Mengersen, and M. Pagano. A geostatistical model for combined analysis of point-level and area-level data using INLA and SPDE. *Spatial Statistics*, 21:27 – 41, 2017. doi: 10.1016/j.spasta.2017.04.006.

Q. Mu, F.A Heinsch, M. Zhao, and S.W Running. *Remote Sensing of Environment*, 111:519–536.

E. Myrvoll-Nilsen, S. Sørbye, H-B. Fredriksen, H. Rue, and M. Rypdal. Statistical estimation of global surface temperature response to forcing under the assumption of temporal scaling. *Earth System Dynamics*, 11:329–345, 04 2020. doi: 10.5194/esd-11-329-2020.

NVE. *Kraftproduksjon*, 2020. https://www.nve.no/energiforsyning/kraftproduksjon/?ref=mainmenu, (accessed April 23, 2020).

T. Opitz, R. Huser, H. Bakka, and H. Rue. Inla goes extreme: Bayesian tail regression for the estimation of high spatio-temporal quantiles. *Extremes*, 21, 02 2018. doi: 10.1007/s10687-018-0324-x.

C. Paciorek and M. Schervish. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17:483–506, 08 2006. doi: 10.1002/env.785.

L. Pannecoucke, M. Le Coz, X. Freulon, and C. de Fouquet. Combining geostatistics and simulations of flow and transport to characterize contamination within the unsaturated zone. *Science of The Total Environment*, 699:134216, 2020. ISSN 0048-9697. doi: https://doi.org/10.1016/j.scitotenv.2019.134216. URL http://www.sciencedirect.com/science/article/pii/S0048969719341932.

M. Plummer. Jags: A program for analysis of bayesian graphical models using gibbs sampling. *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria*, 124, 04 2003.

N. Qiu, X. Chen, Q. Hu, J. Liu, R. Huang, and M. Gao. Hydro-stochastic interpolation coupling with the budyko approach for prediction of mean annual runoff. *Hydrology and Earth System Sciences*, 22(5):2891–2901, 2018. doi: 10.5194/hess-22-2891-2018. URL https://www.hydrol-earth-syst-sci.net/22/2891/2018/.

M. Rivest, D. Marcotte, and P. Pasquier. Hydraulic head field estimation using kriging with an external drift: A way to consider conceptual model information. *Journal of Hydrology*, 361(3):349 – 361, 2008. ISSN 0022-1694. doi: https://doi.org/10.1016/j.jhydrol.2008.08.006.

T. Roksvåg, I. Steinsland, and K. Engeland. A geostatistical two field model that combines point observations and nested areal observations, and quantifies long-term spatial variability – a case study of annual runoff predictions in the voss area. *arXiv:1904.02519v4*, 2020a.

T. Roksvåg, I. Steinsland, and K. Engeland. Estimation of annual runoff by exploiting long-term spatial patterns and short records within a geostatistical framework. *Accepted for publication in Hydrology and Earth System Sciences*, 2020b.

H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 2005.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71:319–392, 2009. doi: 10.1111/j.1467-9868.2008.00700.x.

E. Sauquet. Mapping mean annual river discharges: geostatistical developments for incorporating river network dependencies. *Journal of Hydrology*, 331(1-2):p. 300 – p. 314, 2006. URL `https://hal.archives-ouvertes.fr/hal-00451718`.

E. Sauquet, L. Gottschalk, and E. Lebois. Mapping average annual runoff: A hierarchical approach applying a stochastic interpolation scheme. *Hydrological Sciences Journal*, 45(6):799–815, 2000. doi: 10.1080/02626660009492385.

A. Schmidt and A. Gelfand. A bayesian coregionalization approach to multivariate pollutant data. *Journal of Geophysical Research*, 108, 12 2003. doi: 10.1029/2002JD002905.

J. Sicacha Parada, I. Steinsland, B. Cretois, and J. Borgelt. Accounting for spatial varying sampling effort due to accessibility in citizen science data: A case study of moose in norway. *Spatial Statistics*, page 100446, 04 2020. doi: 10.1016/j.spasta.2020.100446.

D. Simpson, H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, 32:1–28, 2017. doi: 10.1214/16-STS576.

J. O. Skøien, R. Merz, and G. Blöschl. Top-kriging - geostatistics on stream networks. *Hydrology and Earth System Sciences Discussions*, 10:277–287, 2006. doi: 10.5194/hess-10-277-2006.

J. O. Skøien, G. Blöschl, and A. W. Western. Characteristic space scales and timescales in hydrology. *Water Resources Research*, 39, 2003. doi: 10.1029/2002WR001736.

D. Spiegelhalter, A. Thomas, N. Best, and W. Gilks. Bugs: Bayesian inference using gibbs sampling, version 0.5 (version ii). 07 1999.

M. Stein. *Interpolation of spatial data. Some theory for kriging.* Springer series in statistics. Springer, New York, 1999.

S. Su, C. Lei, A. Li, J. Pi, and Z. Cai. Coverage inequality and quality of volunteered geographic features in chinese cities: Analyzing the associated local characteristics using geographically weighted regression. *Applied Geography*, 78:78 – 93, 2017. ISSN 0143-6228. doi: https://doi.org/10.1016/j.apgeog.2016.11.002.

Sveriges Meteorologiska och Hydrologiska Institut. *The HBV Model - Its Structure and Applications.* SMHI reports hydrology. Sveriges Meteorologiska och Hydrologiska Institut, 1992.

T. L. Thorarinsdottir, K. H. Hellton, G. H. Steinbakk, L. Schlichting, and K. Engeland. Bayesian regional flood frequency analysis for large catchments. *Water Resources Research*, 54(9):6929–6947, 2018. doi: 10.1029/2017WR022460.

L. Tierney and J. B Kadane. Accurate approximations for posterior moments and

marginal densities. *Journal of the American Statistical Association*, 81(393):82–86, 1986. doi: 10.1080/01621459.1986.10478240.

W. Tobler. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 1970.

A. Viglione, J. Parajka, M. Rogger, J L. Salinas, G. Laaha, M. Sivapalan, and G Blöschl. Comparative assessment of predictions in ungauged basins - Part 3: Runoff signatures in Austria. *Hydrology and Earth System Sciences Discussions*, 10:449–485, 2013. doi: 10.5194/hessd-10-449-2013.

R. M. Vogel and J. R. Stedinger. Minimum variance streamflow record augmentation procedures. *Water Resources Research*, 21(5):715–723, 1985. doi: 10.1029/WR021i005p00715.

H. Wackernagel. *Multivariate Geostatistics — An Introduction With Applications*. Springer-Verlag, New York, 2nd ed. edition, 1998.

J. Wakefield, G-A. Fuglstad, A. Riebler, J. Godwin, and S. Clark. Estimating under five mortality in space and time in a developing world context. *Statistical Methods in Medical Research*, 28, 11 2017. doi: 10.1177/0962280218767988.

C. Wang, M.A. Puhan, and R. Furrer. Generalized spatial fusion model framework for joint analysis of point and areal data. *Spatial Statistics*, 23:72 – 90, 2018.

P. Whittle. On stationary processes in the plane. *Biometrika*, 41:434–49, 1954.

P. Whittle. Stochastic processes in several dimensions. *Bulletin of the International Statistical Institute*, 40:974–994, 1963.

E. Widén-Nilsson, S. Halldin, and Chong yu Xu. Global water-balance modelling with wasmod-m: Parameter estimation and regionalisation. *Journal of Hydrology*, 340(1): 105 – 118, 2007. doi: https://doi.org/10.1016/j.jhydrol.2007.04.002.

WMO. *International meteorological vocabulary*. 1992.

M.A Wolff, A. Petersen-Øverleir, K. Ødemark, T. Reitan, and R. Brækkan. Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: Results of a Norwegian field study. *Hydrology and Earth System Sciences*, 19, 2015.

Y. Yuan, F. Bachl, F. Lindgren, D. Borchers, J. Illian, S. Buckland, H. Rue, and T. Gerrodette. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *The Annals of Applied Statistics*, 11:2270–2297, 12 2017. doi: 10.1214/17-AOAS1078.

K. Zhang, J. S. Kimball, Q. Mu, L. A. Jones, S. J. Goetz, and S. W. Running. Satellite based analysis of northern ET trends and associated changes in the regional water balance from 1983 to 2005. *Journal of Hydrology*, 379:92 – 110, 2009.

Z. Zrinji and D. H. Burn. Flood frequency analysis for ungauged sites using a region of influence approach. *Journal of Hydrology*, 153(1):1 – 21, 1994. ISSN 0022-1694. doi: https://doi.org/10.1016/0022-1694(94)90184-8.

Abstracts

Paper A

A geostatistical two field model that combines point observations and nested areal observations, and quantifies long-term spatial variability - A case study of annual runoff predictions in the Voss area

Thea Roksvåg, Ingelin Steinsland and Kolbjørn Engeland
(In revision in a statistics journal)

We estimate annual runoff by using a Bayesian geostatistical model for interpolation of hydrological data of different spatial support. That is, streamflow observations from catchments (areal data), and precipitation and evaporation data (point data). The model contains one climatic spatial effect that is common for all years under study, and one year specific spatial effect. Hence, the framework enables a quantification of the spatial variability caused by long-term weather patterns and processes. This can contribute to a better understanding of biases and uncertainties in environmental modeling. The suggested model is tested by predicting 10 years of annual runoff for around Voss in Norway and through a simulation study. We find that on average we benefit from combining point and areal data compared to using only one of the data types, and that the interaction between nested areal data and point data gives a spatial model that takes us beyond smoothing. Another finding is that when climatic effects dominate over annual effects, systematic under- and overestimation of runoff over time can be expected. On the other hand, a dominating climatic spatial effect implies that short records of runoff from an otherwise ungauged catchment can lead to large improvements in the predictions.

Paper B

Estimation of annual runoff by exploiting long-term spatial patterns and short records within a geostatistical framework

Thea Roksvåg, Ingelin Steinsland and Kolbjørn Engeland
Accepted for publication in Hydrology and Earth System Sciences

*In this article, we present a Bayesian geostatistical framework that is particularly suitable for interpolation of hydrological data when the available dataset is sparse and includes both long and short records of runoff. A key feature of the proposed framework is that several years of runoff are modeled simultaneously with two spatial fields: One that is common for all years under study that represents the runoff generation due to long-term (climatic) conditions, and one that is year specific. The climatic spatial field captures how short records of runoff from partially gauged catchments vary relative to longer time series from other catchments, and transfers this information across years. To make the Bayesian model computationally feasible and fast, we use integrated nested Laplace approximations (*INLA*) and the stochastic partial differential equation (*SPDE*) approach to spatial modeling.*

The geostatistical framework is demonstrated by filling in missing values of annual runoff and by predicting mean annual runoff for around 200 catchments in Norway. The predictive performance is compared to Top-Kriging (interpolation method) and simple linear regression (record augmentation method). The results show that if the runoff is driven by processes that are repeated over time (e.g. orographic precipitation patterns), the value of including short records in the suggested model is large. For partially gauged catchments the suggested framework perform better than comparable methods, and one annual observation from the target catchment can lead to a 50 % reduction in RMSE *compared to when no observations are available from the target catchment. We also find that short records safely can be included in the framework regardless of the spatial characteristics of the underlying climate, and down to record lengths of one year.*

Paper C

A geostatistical spatially varying coefficient model for mean annual runoff that incorporates process-based simulations and short records

Thea Roksvåg, Ingelin Steinsland and Kolbjørn Engeland

We present a Bayesian geostatistical model for mean annual runoff predictions that uses the observed streamflow as a response variable and that incorporates the simulations from a process-based hydrological model through a covariate. The regression coefficient of the covariate is modeled as a spatial field such that the relationship between the covariate (simulations from a hydrological model) and the response variable (observed streamflow) is allowed to vary within the study area. Hence, it is a spatially varying coefficient. A preprocessing step for including short records in the modeling is also suggested.

The geostatistical model is demonstrated by predicting mean annual runoff for 1981-2010 for 127 catchments in Norway based on observations from 411 catchments. Simulations from the process-based HBV *model on a 1 km × 1 km grid are used as input. We found that on average the proposed approach outperformed a purely process-based approach (*HBV*) when predicting runoff for ungauged and partially gauged catchments: The reduction in* RMSE *compared to the* HBV *model was 20 % for ungauged catchments and 58 % for partially gauged catchments, where the latter is due to the preprocessing step. For ungauged catchments the proposed framework also outperformed a purely geostatistical method with a 10 % reduction in* RMSE *compared to the geostatistical method. For partially gauged catchments however, a purely geostatistical method performed equally well as the proposed approach. In our evaluation, we also show how the suggested model is able to construct runoff maps (with uncertainty quantification) that preserve the details provided by the original process-based hydrological input model. This is a benefit over purely geostatistical methods that only do spatial smoothing.*

Technical note

An overview of a geostatistical two field model for interpolation of environmental variables

Thea Roksvåg

In this note, we explore the properties of a Bayesian geostatistical two field model for spatio-temporal data. It is suitable for interpolating environmental variables that follow repeated spatial patterns over time and for modeling long-term averages. The two field model was first presented in Roksvåg et al. (2020a,b) where it was used for interpolation of annual runoff. In this note we answer some questions that were considered outside the scope of these articles: For example, we illustrate when the two field model is useful compared to other statistical models for spatial and spatio-temporal data. This is done by simulation experiments and by presenting a new collection of case examples. General model properties such as parameter identifiability are also investigated, and we provide `R` *code for for implementing the model in* `r-inla`*. This represents new contributions compared to the work in Roksvåg et al. (2020a,b).*

Paper A

A geostatistical two field model that combines point observations and nested areal observations, and quantifies long-term spatial variability - A case study of annual runoff predictions in the Voss area

Thea Roksvåg, Ingelin Steinsland and Kolbjørn Engeland

In revision in a statistics journal

A geostatistical two field model that combines point observations and nested areal observations, and quantifies long-term spatial variability - A case study of annual runoff predictions in the Voss area.

Thea Roksvåg[1], Ingelin Steinsland[1] and Kolbjørn Engeland[2]
[1] Department of Mathematical Sciences, NTNU, Norway
[2]The Norwegian Water Resources and Energy Directorate (NVE)

Abstract

We estimate annual runoff by using a Bayesian geostatistical model for interpolation of hydrological data of different spatial support. That is, streamflow observations from catchments (areal data), and precipitation and evaporation data (point data). The model contains one climatic spatial effect that is common for all years under study, and one year specific spatial effect. Hence, the framework enables a quantification of the spatial variability caused by long-term weather patterns and processes. This can contribute to a better understanding of biases and uncertainties in environmental modeling.

The suggested model is tested by predicting 10 years of annual runoff for around Voss in Norway and through a simulation study. We find that on average we benefit from combining point and areal data compared to using only one of the data types, and that the interaction between nested areal data and point data gives a spatial model that takes us beyond smoothing. Another finding is that when climatic effects dominate over annual effects, systematic under- and overestimation of runoff over time can be expected. On the other hand, a dominating climatic spatial effect implies that short records of runoff from an otherwise ungauged catchment can lead to large improvements in the predictions.

1 Introduction

Data related to meteorology, geology and hydrology are often connected to geographical locations. The data are typically linked to point locations, but there are also data observed over an areal unit, e.g. over a crop field, a forest, a grid from a satellite observation or an administrative unit like a country. While point referenced data give information about the process of interest at one specific location, the areal referenced data impose a constraint on the process and/or contain information about aggregated or mean values in a larger area.

For some processes, there exist point data *and* areal data that give information about the same underlying process, and studies show that both observation types should be taken into account when making statistical inference and predictions (Moraga et al., 2017; Wang et al., 2018). There are several challenges connected to simultaneously use data of different spatial support: The data types must be connected to the process of interest in a meaningful way, and expert opinions about the involved measurement uncertainties should be taken into account. In addition, information about how the point and areal data are related to each other is important, such that the observation types can be combined in a mathematically consistent way that preserves basic physical laws (i.e. the conservation of mass and energy).

In this article we consider runoff, which is an example of a process that can be observed through point and areal data. Runoff is defined as the part of the precipitation that flows towards a river on the ground surface (surface runoff) or within the soil (subsurface runoff or interflow) (WMO, 1992). Every point in the landscape contributes to runoff generation, and on an annual scale runoff can be approximated by the estimated point precipitation minus the actual point evaporation at a location of interest (Sauquet et al., 2000). With this interpretation, runoff is a continuous point referenced process in space. However, runoff accumulated over an area is typically observed by measuring the amount of water that flows through the outlet of a stream. The observed value does not primarily provide information about the runoff at the location of the stream outlet: It primarily provides information about the runoff generating process in

the whole drainage area which is called a catchment. Such observations of runoff are therefore areal referenced.

Since most catchments in the world are ungauged (i.e. without runoff observations), a common task for hydrologists is to predict runoff in these catchments. In this article we consider predictions of *annual* runoff which is a key hydrological signature. The annual runoff gives information about the total amount of water available in an area of interest and is fundamental for water resources management, i.e. in the planning of domestic, agricultural, and industrial water supply, and for allocation of water between stakeholders. Annual runoff is also commonly used as a key variable when predicting other runoff properties in ungauged catchments, i.e. low flows and floods (Blöschl et al., 2013). Furthermore, the variability in annual runoff is interesting as it is a key quantity for understanding runoff's sensitivity to driving climatic factors in today's climate, and can be used to make inference about the runoff variability also for future climates.

There are several approaches to predict runoff in ungauged catchments in hydrology, e.g. process-based methods (Beldring et al., 2003; Widén-Nilsson et al., 2007) and geostatistical methods (Gottschalk, 1993; Sauquet et al., 2000; Skøien et al., 2006). In this article, we choose a geostatistical approach. Within the geostatistical framework, runoff predictions in ungauged catchments have typically been done by interpolation of areal referenced runoff data by using Kriging methods (see e.g. Skøien et al. (2006) or Sauquet et al. (2000)). This has shown promising results. In these methods, precipitation data have often been avoided as an information source because these data are known to be uncertain and/or biased (see e.g. Neff (1977), Groisman and Legates (1994) or Wolff et al. (2015)). Evaporation data are even more uncertain: It is seldom observed directly, but derived from meteorological observations and process-based models like in e.g. Mu et al. or Zhang et al. (2009). In spite of the large uncertainties linked to precipitation and evaporation measurements, precipitation and evaporation are the main drivers behind runoff, and it is reasonable to believe that these data sources can contribute to an increased understanding of the runoff generating process if used cleverly. Particularly in areas with few streamflow observations.

Motivated by this, we present a Bayesian geostatistical model for annual

runoff where we in addition to runoff data, use precipitation and evaporation data for spatial interpolation. The suggested model is a Bayesian hierarchical model where the observation likelihood consists of areal referenced runoff observations from catchments and/or point observations of runoff, where the point observations are annual evaporation subtracted from annual precipitation. Informative priors based on expert knowledge are used on the measurement uncertainties to express our doubt on the precipitation and evaporation data, and to put more weight on the runoff observations that are considered more reliable.

The catchments we study in this article are located around Voss in western Norway. Voss is a mountainous area, and the areas west for Voss are among the wettest in Europe with annual precipitation around 3 m/year. This makes Voss flood exposed, and accurate runoff models are of high importance. Voss is also a challenging area when it comes to runoff estimation due to large spatial variability and low stream gauge density. However, there are several precipitation gauges in the area that can be exploited to increase the hydrological understanding. This makes the Voss area a good candidate for performing spatial interpolation of runoff by also including precipitation and evaporation data.

The large annual precipitation in western Norway is mainly caused by the orographic enhancement of frontal precipitation formed around extratropical cyclones. The orographic enhancement is explained by steep mountains that create a topographic barrier for the western wind belt, which transports moist air across the North Atlantic (Stohl et al., 2008). The topography and the elevation differences result in prominent patterns in precipitation and runoff.

Motivated by the strong orographic effect, we include a spatial component in the model that is constant over the years for which we have runoff observations. This represents the spatial variability of runoff caused by climatic conditions in the study area. Furthermore, it is reasonable to assume that not all of the spatial variability can be explained by the climate, and we include an additional spatial effect to describe the annual discrepancy from the climate.

The climatic part of the model is interesting because it let us quantify how

much of the spatial variability that can be explained by long-term effects. Separating long-term spatial variability from year dependent effects can lead to a better understanding of systematic biases and uncertainties that occur in the prediction of environmental variables due to weather patterns and processes that are more or less apparent each year. A consequence of including the climatic component is also that we obtain a model for which it is possible to exploit short records of data: The climatic component captures how the short records vary relatively to longer data series from nearby catchments. This is a valuable property because sparse datasets are common in hydrology. There are several studies on how short records of runoff can be used to estimate different hydrological signatures (Fiering, 1963; Laaha and Blöschl, 2005), but our framework represents a new approach by incorporating the short records into a geostatistical framework where several years of runoff are modeled simultaneously through a climatic spatial field.

Making inference and predictions with geostatistical models often lead to computational challenges due to matrix operations on (dense) covariance matrices, and in our suggested model we have not only one, but two spatial fields. Our solution to the computation challenges is to use the SPDE-approach to spatial modeling from Rue et al. (2009). Rue et al. (2009) utilizes that a Gaussian random field (GRF) with a Matérn covariance function can be expressed as the solution of a stochastic partial differential equation (SPDE). By approximating the solution of the SPDE by using the finite element method (Brenner and Scott, 2008), the involved GRFs can be expressed as Gaussian Markov random fields (GMRFs). The GMRF approximations enable fast simulation and inference (Rue and Held, 2005), and integrated nested Laplace approximations (INLA) can be applied (Rue et al., 2009).

In geostatistical methods used for runoff interpolation it is common to link the involved catchments to point locations in space, not to areas (see e.g. Merz and Blöschl (2005) or Skøien et al. (2003a)). However, interpreting catchment runoff as point referenced can lead to a violation of basic conservation laws: A significant property of catchments is that they are organized into subcatchments, and for annual runoff the water balance must be conserved for all subcatchments. That is, the total amount of annual

runoff in a subcatchment cannot be larger than the total annual runoff in the main catchment. In the Top-Kriging approach developed by Skøien et al. (2006) the nested structure of catchments is taken into account by computing the covariance between two catchments based on the pairwise distance between all the grid nodes in a discretization of the target catchments. This way, information from a subcatchment is weighted more than information from a nearby non-overlapping catchment. The Top-Kriging approach is currently one of the leading interpolation methods for runoff, and has outperformed other methods in predicting several hydrological signatures in Austria (Viglione et al., 2013).

Our model is similar to the Top-Kriging approach by that we consider streamflow observations as areal referenced and compute the covariance between two catchments accordingly. However, our methodology differs from Top-Kriging and other hydrological interpolation methods by using precipitation (point) data in the interpolation framework in addition to nested streamflow (areal) data. As this is an important difference, one of the main objectives of this paper is to:
1) Explore how the runoff predictions in Voss are influenced by the two different observation types (point and areal observations), and assess if the combination of point and areal data can contribute to an increased predictive performance.

Furthermore, the model we suggest ensures that the water balance is preserved for any point in the landscape by defining annual runoff in a catchment as the integral of the point runoff over the catchment's area. Top-Kriging and other geostatistical models don't necessarily provide a full preservation of the water balance. A second objective is therefore to:
2) Show by example how the interaction between point observations and nested areal observations can contribute to improved predictions of annual runoff because the water balance is taken into account.

A geostatistical model that combines point and areal data in the same way as we do already exists in the literature in Moraga et al. (2017). What is new in our model in terms of statistical modeling is the climatic spatial component. A final objective of the paper is thus to:
3) Present a model for which the spatial variability due to long-term spatial

patterns can be quantified, and show how this can be used as a tool for understanding the uncertainty and biases in the modeling of environmental variables, and for exploiting short records of data.

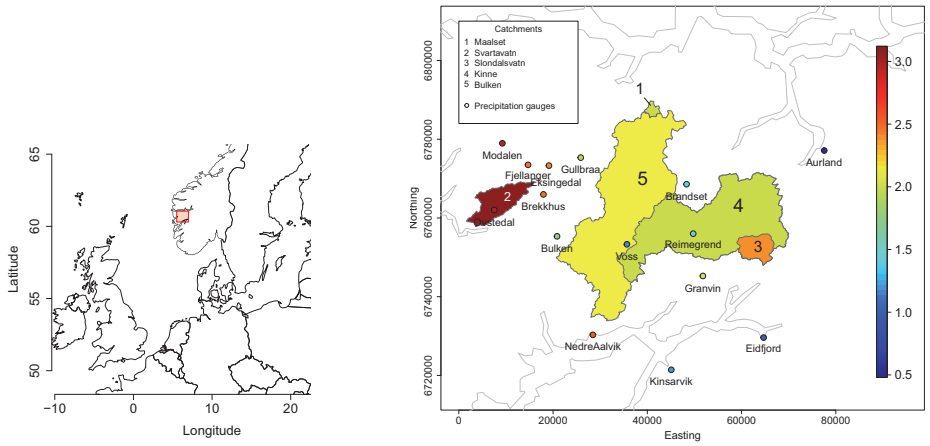
In the section that follows, we present the study area and the available data. Next, we introduce the theoretical background needed to develop the suggested runoff model that is presented in Section 4. In Section 5 the suggested model is fitted to the Voss data. Based on some observation schemes described in Section 5.1, the predictability of annual runoff in Voss is evaluated and discussed. To further demonstrate the value of including a climatic spatial field in the model, a simulation study was carried out. This is presented in Section 6. Finally, our key findings are discussed in Section 7.

2 Study area and data

When modeling hydrological processes on an annual scale, it is common to use the hydrological definition of a year. The basic water balance equation is given as $P = Q + E + S$, where P is precipitation, Q is runoff, E is evapotranspiration and S is the change in stored water (i.e. snow, or groundwater). A hydrological year is defined such that the storage component in the water balance equation can be neglected, i.e. S is much smaller than P and Q. In Norway a hydrological year starts September 1st and ends August 31st, e.g. 1988 begins September 1st 1987 and ends August 31st 1988.

In this analysis, we have runoff data from the hydrological years 1988-2014. The dataset was provided by the Norwegian Water Resources and Energy directorate (NVE) and consists of annual runoff observations from five catchments where three of them are nested (see Figure 1). The unit of the data is m/year and gives the spatial average of the runoff within a catchment. The observations from 1988-1997 are used to make statistical inference, while the observations from 1998-2014 are used as a test set for assessing the model's ability to predict runoff for future years.

The annual runoff data were created by aggregating daily streamflow measurements. The stream gauges that gather the daily observations, don't



(a) The study area is located around Voss in Western Norway.

(b) Mean annual observations [m/year].

Figure 1: Mean annual runoff from 5 catchments and mean annual precipitation minus evaporation (m/year) at 15 precipitation gauges for 1988-1997. Catchment 3 is a subcatchment of Catchment 4 and 5, and Catchment 4 is a subcatchment of Catchment 5. Catchment 1 and Catchment 2 don't overlap with any of the other catchments. The coordinate system in Figure 1b is utm33N.

measure runoff directly: They measure the river's daily stage. Runoff observations are then obtained by using a rating curve that gives the relationship between the stage of the water and the discharge or runoff at a specific point in the stream. The stage-discharge relationship is developed empirically by measuring the discharge across a cross-section of the specific river for a range of stream stages.

Errors in the observed runoff are composed of errors related to the river stage measurement process and errors in the rating curve model. However, on an annual time scale, the river stage measurement errors tend to average out, and the main contribution to errors originates from uncertainties in the rating curve. The dataset provided by NVE includes an estimate of the standard deviation of the observation uncertainty for each (annual) runoff

observation, and the standard deviations are relatively small ranging from 0.65 % to 3.2 % of the corresponding observed value. This information is used to make informative priors for the measurement uncertainties in Section 4.3. We refer to Reitan and Petersen-Øverleir (2009) for details on how the observation (rating curve model) uncertainty is obtained.

In addition to runoff data, we have precipitation data from 15 precipitation gauges. Daily precipitation data were downloaded from `www.eKlima.no` which is a web portal maintained by the Norwegian Meteorological Institute. The observations were aggregated to annual values for the hydrological years 1988-1997. The observed precipitation ranges from 0.55 m/year to 4.6 m/year.

The evaporation data used, originate from the satellite remote sensing-based evapotranspiration algorithm presented in Zhang et al. (2010). The dataset consists of global monthly land surface evapotranspiration with spatial resolution of 1 degree (longitude, latitude). Evaporation data for the locations of the precipitation gauges around Voss were extracted, and monthly values were aggregated to hydrological years (1988-1997). As the spatial resolution of the gridded evaporation dataset is 1 degree and the study area is rather small, the observed annual evaporation within a specific year is the same for almost all of the precipitation gauges. The observed evaporation ranges from 0.23-0.32 m/year with mean 0.25 m/year and standard deviation 0.02 m/year. This means that approximately 12% of the annual precipitation evaporates around Voss, which is a small amount in a global perspective. The observations of evaporation must be considered as approximative estimates of the actual evaporation in the area of interest, with large uncertainties.

Figure 1 shows the 5 catchments where we have measurements of runoff and the locations of the 15 precipitation gauges. Mean annual values for areal referenced runoff and point referenced runoff (precipitation-evaporation) for 1988-1997 are included. We see a spatial pattern with high values of annual runoff in the western part of the study area and low values in the eastern part. This pattern is prominent for all years for which we have data, and indicates that climatic spatial effects dominate over annual spatial effects around Voss.

3 Background

We propose a Latent Gaussian model (LGM) for annual runoff that is computational feasible due to a stochastic partial differential equation (SPDE) formulation of Gaussian random fields (GRFs). In this section we give a brief introduction of these concepts and other relevant background theory and notation for developing and evaluating the model for annual runoff that is presented in Section 4.

3.1 Latent Gaussian Models

In this article we suggest a Latent Gaussian model (LGM) for combining point and areal observations of annual runoff. An LGM can be represented in a hierarchical structure consisting of three levels (see e.g. Gelman et al. (2004)). The first level is the observation likelihood, in this case consisting of two data types $(y_1, ..., y_n)$ and $(z_1, ..., z_m)$. The data are observed with conditional independent likelihood $\Pi_{i=1}^{n}\pi(y_i|q_i, \boldsymbol{\theta_1^y})\Pi_{j=1}^{m}\pi(z_j|Q_j, \boldsymbol{\theta_1^z})$ given two linear predictors q_i and Q_j, and some parameters $(\boldsymbol{\theta_1^y}, \boldsymbol{\theta_1^z})$ which we refer to as hyperparameters. The two linear predictors depend on the same set of latent variables $\boldsymbol{x}$, but connect the data to the latent field differently, through different projection matrices, e.g. $q_i = \boldsymbol{A}_i\boldsymbol{x}$ and $Q_j = \boldsymbol{B}_j\boldsymbol{x}$. Here, $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices that link elements in the latent field to the observations, and $\boldsymbol{A}_i$ and $\boldsymbol{B}_j$ denote row number i and j of the two matrices. The second level of the LGM is formed by the prior of the latent field $\boldsymbol{x}$ and is on the form $\pi(\boldsymbol{x}|\boldsymbol{\theta_2}) \sim \mathcal{N}(\boldsymbol{\mu}(\boldsymbol{\theta_2}), \boldsymbol{\Sigma}(\boldsymbol{\theta_2}))$, i.e. it is Gaussian conditioned on some hyperparameters $\boldsymbol{\theta_2}$. The third level is given by $\pi(\boldsymbol{\theta})$ which is the prior distribution of the hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta_1^y}, \boldsymbol{\theta_1^z}, \boldsymbol{\theta_2})$.

3.2 Gaussian random fields

We use Gaussian random fields (GRFs) to model the spatial variability of annual runoff. A continuous field $\{x(\boldsymbol{u}); \boldsymbol{u} \in \mathcal{D}\}$ defined on a spatial domain $\mathcal{D} \in \mathcal{R}^2$ is a GRF if for any collection of locations $\boldsymbol{u}_1, ..., \boldsymbol{u}_n \in \mathcal{D}$ the vector $(x(\boldsymbol{u}_1), ..., x(\boldsymbol{u}_n))$ follows a multivariate normal distribution

(Cressie, 1993), i.e. $(x(\boldsymbol{u}_1), ..., x(\boldsymbol{u}_n)) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The covariance matrix $\boldsymbol{\Sigma}$ defines the dependency structure in the spatial domain, and can be constructed from a covariance function $C(\boldsymbol{u}_i, \boldsymbol{u}_j)$. Furthermore, the dependency structure for a spatial process is often characterized by two parameters: The marginal variance σ^2 and the range ρ. The marginal variance gives information about the spatial variability of the process of interest, while the range gives information about how the correlation between the process at two locations decays with distance. If the range and marginal variance are constant over the spatial domain, we have a stationary GRF.

One popular choice of covariance function is the Matérn covariance function which is given by

$$C(\boldsymbol{u}_i, \boldsymbol{u}_j) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa||\boldsymbol{u}_j - \boldsymbol{u}_i||)^\nu K_\nu(\kappa||\boldsymbol{u}_j - \boldsymbol{u}_i||), \tag{1}$$

where $||\boldsymbol{u_j} - \boldsymbol{u_i}||$ is the Euclidean distance between two locations $\boldsymbol{u_i}, \boldsymbol{u_j} \in \mathcal{R}^d$, K_ν is the modified Bessel function of the second kind and order $\nu > 0$, $\Gamma(\cdot)$ is the gamma function and σ^2 is the marginal variance (Guttorp and Gneiting, 2006). The parameter κ is the scale parameter, and it can be shown empirically that the spatial range can be expressed as $\rho = \sqrt{8\nu}/\kappa$, where ρ is defined as the distance where the spatial correlation between two locations has dropped to 0.1 (Lindgren et al., 2011). Using a Matérn GRF is convenient because it makes it possible to apply the SPDE approach to spatial modeling which is outlined in the next subsection.

3.3 The SPDE approach to spatial modeling

Making statistical inference and predictions on models including GRFs involve matrix operations on the covariance matrix $\boldsymbol{\Sigma}$. This can lead to computational challenges if the covariance matrix is dense. In this paper, we suggest a model for annual runoff that includes not only one, but two GRFs. Consequently, some simplifications have to be done to make the model computationally feasible. To achieve this, we use that the exact solution of the SPDE

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}}\tau x(\boldsymbol{u}) = \mathcal{W}(\boldsymbol{u}), \quad \boldsymbol{u} \in \mathcal{R}^d, \quad \kappa > 0, \quad \nu > 0, \tag{2}$$

is a Gaussian random field with Matérn covariance function. Here, $\mathcal{W}(\cdot)$ is spatial Gaussian white noise, Δ is the Laplacian, α is a smoothness parameter, κ is the scale parameter in Equation (1), d is the dimension of the spatial domain and τ is a parameter controlling the variance. The parameters of the Matérn covariance function in Equation (1) is linked to the SPDE through

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)(4\pi)^{d/2}\kappa^{2\nu}\tau^2}; \qquad \nu = \alpha - d/2,$$

where we will use that $d = 2$ and set $\alpha = 2$, such that ν is fixed to $\nu = 1$. The parameter ν is fixed because it is difficult to identify from data, and $\alpha = 2$, $\nu = 1$ are commonly used values for these parameters (Ingebrigtsen et al., 2014; Blangiardo and Cameletti, 2015).

The link between the above SPDE and the Matérn GRF, which was developed by Whittle (1954, 1963), is used by Lindgren et al. (2011) to show that a GRF can be approximated by a Gaussian Markov random field (GMRF). This is done by solving the SPDE in Equation (2) by the finite element method (FEM) (see e.g Brenner and Scott (2008)). A GMRF is simply a multivariate Gaussian vector that is parametrized by the precision matrix $\boldsymbol{Q}$, which is the inverse $\boldsymbol{\Sigma}^{-1}$ of the covariance matrix. The term GMRF is most used for Gaussian processes with sparse precision matrices, i.e. matrices that contain many zero elements. The zero elements correspond to Markov properties, in this case conditional independence between locations in the spatial domain. It is convenient to work with GMRFs because there exist computationally efficient algorithms for sparse matrix operations (Rue and Held, 2005). Hence, through the SPDE approach from Lindgren et al. (2011) a GRF with a dense precision matrix can be replaced by a GMRF with a sparser precision matrix with computational benefits.

3.4 PC priors

As we use a Bayesian approach, the hyperparameters $\boldsymbol{\theta}$ from Section 3.1 must be given prior distributions. For the majority of the hyperparameters we use penalized complexity (PC) priors. PC priors are proper prior distributions developed by Simpson et al. (2017). The main idea behind PC

priors is to penalize the increased complexity induced by deviating from a simple base model. One of the goals is to avoid overfitting.

The PC prior for the precision τ of a Gaussian effect $\mathcal{N}(0, \tau^{-1})$ has density

$$\pi(\tau) = \frac{\lambda}{2}\tau^{-3/2}\exp(-\lambda\tau^{-1/2}), \qquad \tau > 0, \quad \lambda > 0, \tag{3}$$

where λ is a parameter that determines the penalty of deviating from the base model. The parameter λ can be specified through a quantile u and probability α by $\text{Prob}(1/\sqrt{\tau} > u) = \alpha$, where $u > 0$, $0 < \alpha < 1$ and $\lambda = -ln(\alpha)/u$. Here, $1/\sqrt{\tau}$ is the standard deviation of the Gaussian distribution.

As the range and the marginal variance are easier to interpret than the Matérn covariance function parameters κ and τ in Equation (1), we parametrize our model through ρ and σ. For ρ and σ we use the prior suggested in Fuglstad et al. (2019). This is a joint prior for the spatial range ρ and the marginal variance σ constructed from PC priors. The joint prior can be specified through

$$\text{Prob}(\rho < u_\rho) = \alpha_\rho; \qquad\qquad \text{Prob}(\sigma > u_\sigma) = \alpha_\sigma,$$

where u_ρ, u_σ, α_ρ and α_σ are quantiles and probabilities that must be determined.

3.5 Evaluating the predictive performance

To evaluate the predictive performance of the suggested runoff model, we use two criteria: The first criterion is the root mean squared error (RMSE). The RMSE measures the difference between a point prediction $\hat{y}_i$ and the observed value y_i by

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

where n is the total number of pairs of predictions and observations. We use the posterior mean as a point prediction when computing the RMSE.

The second criterion is the continuous ranked probability score (CRPS). The CRPS is defined as

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(s) - 1\{y \leq s\})^2 ds,$$

where F is the predictive cumulative distribution and y is the observed value (Gneiting and Raftery, 2007). The CRPS takes the whole posterior predictive distribution into account, not only the posterior mean or median, and is penalized if the observed value falls outside the posterior predictive distribution. Both the RMSE and the CRPS are negatively oriented, and a smaller value indicates a better prediction.

3.6 Interpolation by using Top-Kriging

The focus of this article is mainly on highlighting properties of the suggested point and areal runoff model. However, we also compare some of our results to the predictive performance of Top-Kriging. Top-Kriging (Skøien et al., 2006) is one of the leading methods for runoff interpolation. It is a Kriging approach (Cressie, 1993) where it is assumed that the variable of interest can be modeled as a GRF. A prediction of the target variable at an unobserved location is given by a weighted sum of the available observations, and the interpolation weights are estimated by finding the so-called best linear unbiased estimator (BLUE).

In the computation of the interpolation weights, the Top-Kriging approach calculates the covariance between two catchments based on the distance between all the grid nodes in a discretization of the involved catchments. As a consequence, a subcatchment get a higher Kriging weight than a nearby, non-overlapping catchment. This is different from other Kriging approaches traditionally used in hydrology, for which streamflow observations have been treated as point referenced (see e.g. Merz and Blöschl (2005); Skøien et al. (2003a); Adamowski and Bocci (2001)).

While the suggested Bayesian approach for runoff interpolation supports both areal and point observations, Top-Kriging only considers runoff (areal) data. Furthermore, Top-Kriging estimates the covariance (or variogram) empirically, while we take a fully Bayesian approach where the latent field

and the parameters are estimated jointly. Another main difference is that Top-Kriging treats each year of runoff data separately, while we can model several years of runoff simultaneously through our two field model.

4 Statistical Model for Annual Runoff

In this section we present the proposed LGM for annual runoff which is suitable for combining observations of different spatial support and that has a climatic spatial field that let us quantify long-term spatial variability.

4.1 Spatial model for runoff

Let the spatial process $\{q_j(\boldsymbol{u}) : \boldsymbol{u} \in \mathcal{D}\}$ denote the runoff generating process at a point location $\boldsymbol{u}$ in the spatial domain $\mathcal{D} \in \mathcal{R}^2$ in year j. The true runoff generation at point location $\boldsymbol{u}$ is modeled as

$$q_j(\boldsymbol{u}) = \beta_c + c(\boldsymbol{u}) + \beta_j + x_j(\boldsymbol{u}), \quad j = 1, ..., r. \tag{4}$$

Here, the parameter β_c is an intercept common for all years $j = 1, ..., r$, while $c(\boldsymbol{u})$ is a spatial effect common for all years. These two model components represent the runoff generation caused by the climate in the study area. Mark that the term climate here covers all long-term effects, i.e. both long-term weather patterns *and* patterns that are repeated due to catchment characteristics. Further, we include a year specific intercept β_j and a year specific spatial effect $x_j(\boldsymbol{u})$ for $j = 1, ..r$ to model the runoff generation due to the annual discrepancy from the climate. Both spatial effects $c(\boldsymbol{u})$ and $x_j(\boldsymbol{u})$ are modeled as GRFs with zero mean and Matérn covariance functions given the model parameters; $c(\boldsymbol{u})$ with range parameter ρ_c and marginal variance σ_c^2, and $x_j(\boldsymbol{u})$ with range ρ_x and marginal variance σ_x^2. The spatial fields $x_j(\boldsymbol{s})$, j=1,...,r, are assumed to be independent realizations, or replicates of the same underlying GRF. The same applies for the year specific intercepts β_j which are assumed to be independent and identically distributed as $\mathcal{N}(0, \tau_\beta^{-1})$ given the parameter τ_β with $\beta_1, ..., \beta_r$ being independent realizations of this Gaussian distribution.

The true mean runoff generated inside a catchment $\mathcal{A}$ in year j can be expressed as

$$Q_j(\mathcal{A}) = \frac{1}{|\mathcal{A}|} \int_{\boldsymbol{u} \in \mathcal{A}} q_j(\boldsymbol{u}) d\boldsymbol{u}, \quad j = 1, ..., r, \tag{5}$$

where $|\mathcal{A}|$ is the area of catchment $\mathcal{A}$. By interpreting catchment runoff as an integral of point referenced runoff $q_j(\boldsymbol{u})$, we obtain a mathematically consistent model where the water balance is conserved for any point in the landscape.

4.2 Observation model

Annual precipitation and evaporation are observed at n locations $\boldsymbol{u}_i \in \mathcal{D}$ for $i = 1, ..n$ and for r years $j = 1, ..r$. The observed annual runoff generation at point location $\boldsymbol{u}_i$, year j, is modeled as the difference between the observed annual precipitation p_{ij} and annual evaporation e_{ij},

$$y_{ij} = p_{ij} - e_{ij} = q_j(\boldsymbol{u}_i) + \epsilon_{ij}^y \quad i = 1, ..., n; \quad j = 1, ..., r, \tag{6}$$

where $q_j(\boldsymbol{u}_i)$ is the true annual point runoff from Equation (4). The error terms ϵ_{ij}^y are independent and identically distributed as $\mathcal{N}(0, s_{ij}^y \cdot \tau_y^{-1})$ and independent of the other model components. The measurement uncertainties for precipitation and evaporation are assumed to increase with the magnitude of the observed value, and we want to include this assumption in the model. This is done by scaling the precision parameter of the error terms τ_y with a fixed factor s_{ij}^y, that is further described in Section 4.3.

Runoff at catchment level is observed through streamflow data from K catchments denoted $\mathcal{A}_1, ..., \mathcal{A}_K$ for r years denoted $j = 1, .., r$. We use the following model for the annual runoff observed in catchment $\mathcal{A}_k$ in year j

$$z_{kj} = Q_j(\mathcal{A}_k) + \epsilon_{kj}^z \quad k = 1, ..., K; j = 1, ..., r, \tag{7}$$

where $Q_j(\mathcal{A}_k)$ is the true annual areal runoff from Equation (5). The measurement errors ϵ_{kj}^z are independent and identically distributed as $\mathcal{N}(0, s_{kj}^z \cdot \tau_z^{-1})$ and independent of the other model components. As for the

point referenced observations, the precision parameter of the error terms τ_z is scaled with a fixed factor s^z_{kj} that is further described in the next subsection. This way the uncertainty estimates that the data provider NVE has for each annual observation can be included in the modeling.

In Equation (7) the variable $Q(\mathcal{A})$ defines an areal representation of the annual runoff in catchment $\mathcal{A}_k$. Hence, through the likelihood, the annual runoff in catchment area $\mathcal{A}_k$ is constrained to be close to the actually observed value (with some uncertainty).

So far we have defined the observation likelihoods for the point and areal observations separately. To construct a joint model for point and areal runoff, we multiply the likelihoods defined in Equation (6) and (7) together as described in Section 3.1. This is done for all n precipitation gauge locations $i = 1, ..n$, for all catchments $k = 1, ..K$ and for all years $j = 1, ..r$ such that we obtain a model that simultaneously models several years of runoff. Different years are linked together through the climatic part of the model $c(\boldsymbol{u}) + \beta_c$ from Equation (4).

4.3 Prior distributions

In the suggested model for annual runoff there are 8 parameters (τ_y, τ_z,ρ_c, ρ_x, σ_c, σ_x, β_c,τ_β) that must be given prior distributions. We start by formulating priors for the measurement errors for the point and areal observations.

The variance of the measurement error of the point referenced observation from precipitation gauge i, year j, is given by $s^y_{ij}\tau_y^{-1}$ where τ_y is a hyperparameter and s^y_{ij} is a deterministic value that scales the variance based on expert opinions from NVE about the measurement errors for precipitation and evaporation.

The precipitation data are obtained by observing the amount of water or snow that falls into a bucket, but the buckets often fail to catch a large proportion of the actual precipitation, particularly for windy snow events (Neff, 1977; Groisman and Legates, 1994; Wolff et al., 2015). Based on this and recommendations from NVE, the standard deviation of the observation uncertainty for precipitation is assumed to be 10% of the observed value

p_{ij}. The evaporation data are obtained from satellite observations and process-models, and are more uncertain than the precipitation data. We assume that the standard deviation for evaporation is 20% of the observed value e_{ij}. The prior knowledge about the point data is used to specify the scale s_{ij}^y for the point observation y_{ij} at location i and year j as follows

$$\begin{aligned} s_{ij}^y &= \text{Var}(y_{ij}) = \text{Var}(p_{ij} - e_{ij}) \\ &= \text{Var}(p_{ij}) + \text{Var}(e_{ij}) - 2 \cdot \text{Cov}(p_{ij}, e_{ij}) \\ &= (0.1p_{ij})^2 + (0.2e_{ij})^2 - 2 \cdot \text{Cov}(p_{ij}, e_{ij}). \end{aligned}$$

Here, the covariance between the observed precipitation and evaporation is estimated by

$$\text{Cov}(p_{ij}, e_{ij}) \ = \sqrt{\text{Var}(p_{ij})} \cdot \sqrt{\text{Var}(e_{ij})} \cdot \text{Cor}\{(p_{i1}, ..., p_{ir}), (e_{i1}, ..., e_{ir})\},$$

where $Cor\{\cdot, \cdot\}$ is the Pearson correlation between all available observations of precipitation and evaporation at precipitation gauge i. Further, we assign the precision τ_y the PC prior from Equation (3) with $\alpha = 0.1$ and $u = 1.5$. With this prior, a prior 95 % credible interval for the standard deviation $\sqrt{s_{ij}^y \tau_y^{-1}}$ of the measurement error for point runoff becomes around (0.002-30)% of the corresponding observed value y_{ij}. This interval corresponds well to what NVE knows about the measurement uncertainty for precipitation and evaporation.

The same approach is used to make a prior for the variance of the measurement error for the areal referenced observations z_{kj}. The precision τ_z is given a PC prior with $\alpha = 0.1$ and $u = 1.5$, while the scale s_{kj}^z for catchment k, year j is given by

$$s_{kj}^z = \text{Var}(z_{kj}). \tag{8}$$

For the streamflow data, information about the variance of the observations is directly available through the dataset provided by NVE. These data are inserted into Equation (8). With the suggested prior, a prior 95 % credible interval for the standard deviation $\sqrt{s_{kj}^z \tau_z^{-1}}$ of an areal observation, is approximately (0.002,4.0) % of the corresponding observed value z_{kj}. This is an informative prior that just covers the range of values suggested by

NVE. We have chosen a low prior standard deviation in order to try to put more weight on the runoff observations than to the point observations. There are only 5 areal observations available for each year in the dataset, but 15 point observations, and the aim is to avoid that the more unreliable point data dominate over the areal data.

For the spatial ranges and the marginal variances of the spatial fields $x_j(\boldsymbol{u})$ and $c(\boldsymbol{u})$, the joint PC prior from Fuglstad et al. (2019) is used. The PC priors for σ_x, ρ_x, σ_c and ρ_c are specified through the following probabilities and quantiles:

$$\text{Prob}(\rho_x < 10 \text{ km}) = 0.1, \;\; \text{Prob}(\sigma_x > 2 \text{ m/year}) = 0.1,$$
$$\text{Prob}(\rho_c < 10 \text{ km}) = 0.1, \;\; \text{Prob}(\sigma_c > 2 \text{ m/year}) = 0.1.$$

The percentages and quantiles are chosen based on expert knowledge about the spatial variability in the area. The study area is approximately 80 km $\times$ 80 km, and it is reasonable to assume that there is a correlation larger than 0.1 between two locations that are less than 10 km apart. Furthermore, the spatial variability in the study area is large, and we can observe runoff values from 0.8 m/year to 3.2 m/year within the same year. However, it is reasonable to assume that the marginal standard deviation of the runoff generating process does not exceed 2 m/year. The parameters of the climatic GRF $c(\boldsymbol{u})$ and the annual GRF $x_j(\boldsymbol{u})$ are given the same prior as it is difficult to identify if the spatial variability mainly comes from climatic processes or from annual variations. We also want the data to determine which of the two effects that dominates in the study area.

As described in Section 4.1, the year specific intercept β_j has prior $\mathcal{N}(0, \tau_\beta^{-1})$ for all years $j = 1, ..r$. Its precision τ_β is given the PC prior from Equation (3) with $u = 10$ and $\alpha = 0.2$. This is a weakly informative wide prior with a prior 95% interval (0.002,40.5) m/year for the standard deviation $\sqrt{\tau_\beta^{-1}}$ of β_j. Finally, the climatic intercept β_c is given a normal prior, $\beta_c \sim \mathcal{N}(2, 0.5^2)$. This gives a prior 95% credible interval of (1.0,3.0) m/year for β_c which covers all reasonable mean values of annual runoff around Voss.

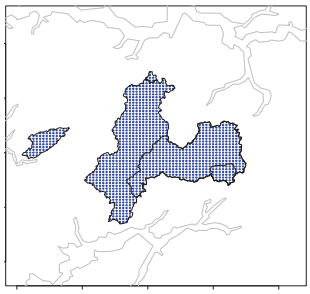
4.4 Inference

In order to make the model computationally feasible, some simplifications of the suggested model are necessary. In Section 4.1 the annual runoff for a catchment $\mathcal{A}_k$ was modeled as the integral of point referenced runoff over the catchment area. In practice, the integral in Equation (5) is calculated by a finite sum over a discretization of the target catchment. More specifically, let $\mathcal{L}_k$ denote the discretization of catchment $\mathcal{A}_k$. The total annual runoff in catchment $\mathcal{A}_k$ in year j is approximated by

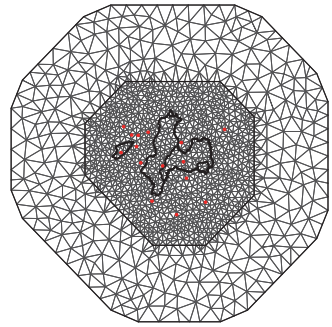
$$Q_j(\mathcal{A}_k) = \frac{1}{N_k} \sum_{\boldsymbol{u} \in \mathcal{L}_k} q_j(\boldsymbol{u}), \tag{9}$$

where N_k is the total number of grid nodes in $\mathcal{L}_k$ and $q_j(\boldsymbol{u})$ is the point runoff at grid node $\boldsymbol{u} \in \mathcal{L}_k$. It is important that a subcatchment shares grid nodes with the main catchment in order to preserve the water balance. The discretization used in this analysis has 1 km spacing and is shown in Figure 2a.

The model suggested for annual runoff, is a latent Gaussian model with the structure described in Section 3.1. Modeling annual runoff as a LGM is convenient because it allows us to use integrated nested Laplace approximations (INLA) to make inference and predictions. INLA can be used for making Bayesian inference on LGMs and is a faster alternative to MCMC algorithms (Gamerman and Lopes, 2006). The approach is based on approximating the marginal distributions by using Laplace or other analytic approximations, and on numerical integration schemes. The main computational tool is the sparse matrix calculations described in Rue and Held (2005), such that in order to work fast, the latent field of the LGM should be a GMRF with a sparse precision matrix. In our case, sparsity is obtained by using the SPDE approach from Section 3.3 to approximate the GRFs $x_j(\boldsymbol{u})$ and $c(\boldsymbol{u})$ by GMRFs. This is done through the finite element method (FEM), and the triangulation used for FEM is shown in Figure 2b. In order to obtain accurate approximations of the underlying two GRFs, this triangular mesh must be dense enough to capture the rapid spatial variability of annual runoff around Voss. If the mesh is too coarse, unrealistic results such as negative runoff can occur, or we can get into numerical problems.



(a) Discretization of catchments used to model areal runoff. We use a regular grid with 1 km spacing.



(b) Triangulation of the spatial domain used for FEM. The red points are the locations of the precipitation gauges.

Figure 2: Discretization and triangular mesh used to make the model computationally feasible.

The R-package `r-inla` was used to make inference and predictions for the suggested model. This package provides a user-friendly interface for applying INLA and the SPDE approach to spatial modeling without requiring that the user has deep knowledge about SPDEs. See `r-inla.org` or Blangiardo and Cameletti (2015) and Krainski et al. (2018) for tutorials and examples. In particular, Moraga et al. (2017) is recommended for a description of how a model with point and areal data can be implemented in `r-inla`.

5 Case study of annual runoff in Voss

The model presented in Section 4 is used to explore the predictability of annual runoff in the Voss area. Recall that the main goals are to investigate how the predictions are affected by the two different observation types (point and areal data), to demonstrate how the water balance considerations can be beneficial, and to explore the properties of the climatic part of the model. To address this, we perform four tests that are inspired

by common applications in hydrology. These are presented in the next subsection. In Section 5.2 the results from the tests are presented and discussed.

5.1 Model evaluation

To explore how the two different observation types influence the predictions of annual runoff around Voss, we compare three observation designs: An observation design where only point referenced observations are included in the likelihood (P), an observation design where only areal referenced observations are included in the likelihood (A) and an observation design where all available observations are included in the likelihood ($P+A$). Recall that using only areal observations (A) corresponds to what typically has been done in hydrological applications, and we want to investigate if we can improve the predictability of runoff by also including point observations in the likelihood ($P+A$). Including P as an observation design gives information about what influence the point data have on the predictions. The three observation designs are evaluated according to four tests that are described as follows:

T1 - Inference: The model from Section 4 is fitted to all available observations between 1988 and 1997 from Figure 1. This is done for P, A and $P + A$, such that we get information about how the different observation types affect the posterior estimates of the parameters.

T2 -Spatial predictions in ungauged catchments: In hydrological applications, the main interest is on estimating runoff at catchment level. Motivated by this, we perform spatial predictions of annual runoff for each of the five catchments $\mathcal{A}_1, ..., \mathcal{A}_5$ by leave-one-out-cross-validation for P, A and $P + A$. That is, data from the target catchment are left out and the catchment of interest is treated as ungauged. Runoff predictions are done for the target catchment for 1988-1997 and are based on observations from the remaining 4 catchments and/or point data from 1988-1997. The predictive performance is assessed by computing the RMSE and CRPS for each catchment based on the 10 years of predictions.

In **T2**, we also compare our results to the Top-Kriging approach described in Section 3.6. For Top-Kriging, we fit the default covariance function (or variogram) from the R package `rtop`. This is a multiplication of a modified exponential and fractal variogram model (Skøien et al., 2006). Recall that Top-Kriging only supports areal referenced (runoff) observations.

T3u- Future predictions in ungauged catchments: In **T2** we estimate the runoff that was generated in ungauged catchments in the past. However, quantifying the annual runoff we can expect in the future is more interesting for most hydrological applications. In **T3u** we therefore estimate annual runoff for a future year, i.e. for a year for which there are no observations of runoff, precipitation or evaporation. For an unobserved year ($j > 10$) the posterior means of the year specific effects β_j and $x_j(\boldsymbol{u})$ are zero. Thus, the posterior predicted future runoff is given by the posterior means of the climatic components β_c and $c(\boldsymbol{u})$. However, all four model components as well as the observation uncertainty contribute to the predictive uncertainty.

In **T3u** the catchment of interest is treated as ungauged and left out of the dataset, and we use the remaining observations from 1988-1997 to predict annual runoff for 1998-2014. This is done for catchment $\mathcal{A}_1, ..., \mathcal{A}_5$ in turn. The predictive performance is evaluated by computing the RMSE and CRPS for predictions of runoff for each of the 5 catchments for 17 future years. The average RMSE and CRPS over the 5 catchments are used as summary scores. As the posterior mean for an unobserved year is given by the posterior mean of the climatic effects β_c and $c(\boldsymbol{u})$, this test lets us quantify the climatology in the study area.

T3g - Future predictions in partially gauged catchments: We predict annual runoff in catchment $\mathcal{A}_1, ..., \mathcal{A}_5$ for a future year as in **T3u**. However, we allow the observation likelihood to contain 1 to 10 annual runoff observations from the catchment in which we want to predict runoff. This way, we assess the model's ability to exploit short records of runoff, which is a property enabled by the climatic component of the model. We denote this test **T3g**, for gauged, as opposed to **T3u** for ungauged.

The test is carried out by drawing i observations between 1988-1997 randomly from the target catchment. Next, these observations are used together with the other point and/or areal observations of P, A and $P + A$ from 1988-1997 to predict the annual runoff in 1998-2014 for this particular catchment. As the experimental results might depend on which runoff observations we pick from the target catchment, the experiment is repeated 10 times such that different observations are included for each experiment.

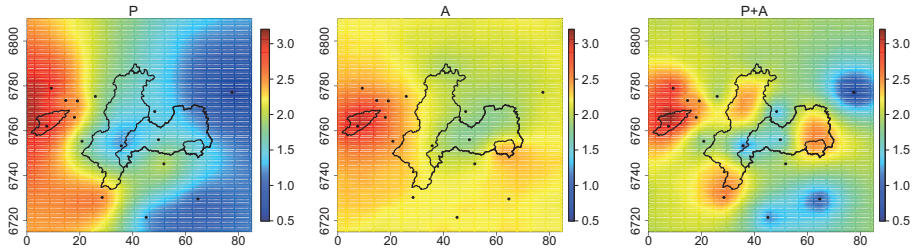
The above procedure is carried out with an increasing number of years included in the short record, i.e. for $i \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. The predictive performance is then evaluated for each i by computing the RMSE and CRPS for each catchment $\mathcal{A}_1, ..., \mathcal{A}_5$ based on 17 years of future predictions. The average RMSE and CRPS over 5 catchments and 10 experiments are reported as summary scores.

For our experiments we use the posterior mean as the predicted value when computing the RMSE. Furthermore, when evaluating the CRPS and when computing the coverage of the predictions, we assume that the posterior distributions are Gaussian with mean given by the posterior mean and standard deviation given by the posterior standard deviation. In the posterior standard deviation, we take the measurement uncertainty given by $s_{kj}^z \tau_z^{-1}$ into account, in addition to the uncertainty of the model components of the linear predictor in Equation (4). The Gaussian distribution should be a good approximation for the resulting posterior distributions as they typically are symmetric with neither particularly short or long tails.

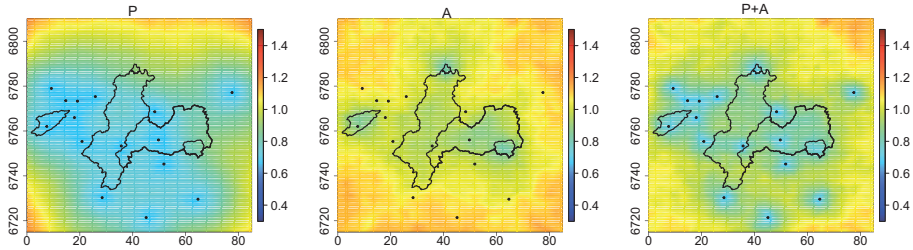
5.2 Results from the case study

We now present the results from the case study for our four tests **T1**, **T2**, **T3u** and **T3g** in turn.

Table 1 shows the posterior medians and the 0.025 and 0.975 quantiles for the hyperparameters for P (point observations), A (areal observations) and $P + A$ (point and areal observations) when all respective available observations from 1988-1997 are used to make inference (**T1**). In general, P gives lower runoff values with a posterior median of the climatic intercept



(a) Posterior mean [m/year].



(b) Posterior standard deviation [m/year].

Figure 3: Posterior mean and standard deviation for annual runoff for a future unobserved year when all available observations of point observations (P, left), areal (A, middle) and both point and areal observations ($P + A$, right) from 1988-1997 are used, i.e. all catchments are treated as gauged for A and $P + A$, but ungauged for P (test **T1**).

β_c equal to 1.87 m/year compared to A giving β_c equal to 2.21 m/year. Furthermore, the posterior median of the marginal standard deviation of the climatic GRF σ_c is considerably larger for P with $\sigma_c = 0.97$ m/year compared to A and $P + A$ which give posterior medians 0.63 m/year and 0.76 m/year respectively. The posterior median of the range of the climatic GRF ρ_c is also larger for P with 70 km compared to values around 20 km for A and $P + A$.

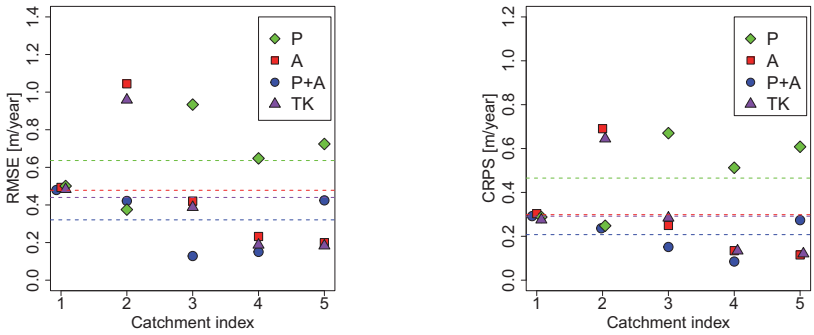
The spatial runoff patterns corresponding to these parameter values are shown in Figure 3. These figures show the posterior mean and standard deviation for runoff for an unobserved, future year. We see that larger

Table 1: Posterior median (0.025 quantile, 0.975 quantile) when all available point (P), areal (A) and both point and areal (P+A) referenced observations from 1988-1997 are used for making inference (test **T1**). The precision parameters are transformed to standard deviations to make them more interpretable. Recall that the posterior estimates of the standard deviations $1/\sqrt{\tau_y}$ and $1/\sqrt{\tau_z}$ of the measurement uncertainties are multiplied with the root of the unitless scales from Section 4.3 in order to obtain the final posterior observation uncertainty with unit [m/year].

Parameter [unit]	Posterior median (0.025 quantile, 0.975 quantile)		
	P	A	P+A
ρ_x [km]	236 (148, 379)	104 (32, 262)	102 (41, 249)
σ_x [m/year]	0.27 (0.20, 0.38)	0.34 (0.18, 0.56)	0.29 (0.19, 0.44)
ρ_c [km]	70 (30, 180)	25 (9, 74)	20 (9, 46)
σ_c [m/year]	0.97 (0.56, 1.79)	0.63 (0.34, 1.34)	0.76 (0.53, 1.1)
β_c [m/year]	1.87 (1.13, 2.68)	2.21 (1.57, 2.82)	1.96 (1.40, 2.50)
$1/\sqrt{\tau_y}$ [m/year]	0.48 (0.40, 0.57)	$\times$	0.37 (0.22, 0.54)
$1/\sqrt{\tau_z}$ [m/year]	$\times$	3.6 (2.3, 5.1)	5.3 (3.8,6.8)
$1/\sqrt{\tau_\beta}$ [m/year]	0.26 (0.01, 0.78)	0.61 (0.31,1.0)	0.48 (0.24, 0.75)

values for ρ_c and σ_c lead to a more prominent spatial pattern for P with large runoff values in the western part of the study area and lower values in the eastern part. A high climatic range ρ_c also leads to a reduction of the posterior predictive uncertainty in a larger part of the study area for P, as can be seen in Figure 3b. The maps show that the choice of observation scheme (P, A or $P + A$) has a large impact on the resulting predictions of annual runoff in terms of posterior mean and/or posterior standard deviation.

In **T2** we perform spatial predictions of annual runoff in 1988-1997 for a catchment that is left out of the dataset. The predictive performance for spatial predictions is summarized in Figure 4. For four out of five catchments, $P + A$ gives the lowest RMSE and CRPS, or a RMSE and CRPS that is approximately as for A, P or Top-Kriging (TK). We see that the Top-Kriging approach performs similar to A, which is reasonable as Top-Kriging only considers areal observations and uses a similar interpretation of covariance as our suggested model.



(a) RMSE. (b) CRPS.

Figure 4: Predictive performance for spatial predictions of runoff in 1988-1997 when the target catchment is treated as ungauged (test **T2**) for P, A and P+A. Here, we have also included results from the reference method Top-Kriging (TK) that only considers areal observations. Dashed lines mark the average performance over all catchments.

In Figure 4 we particularly highlight Catchment 3 because it provides an example of how the water balance properties of the model can be beneficial. Figure 4 shows that for Catchment 3, P gives a RMSE around 0.9, while A gives a RMSE around 0.4. Considering the posterior prediction intervals for Catchment 3 in Figure 5, we see that P leads to an underestimation of the annual runoff. This can be explained by looking at the observations in Figure 1: The point observations close to Catchment 3 all have mean values lower than the true mean annual runoff in this catchment. Next, considering the results for the areal observations (A), Figure 5 shows that also these lead to an underestimation of Catchment 3's runoff. Intuitively, we would thus expect that combining P and A would result in underestimation. Instead, we get a large improvement in the predictions in Figure 5 when P and A are combined, with a RMSE around 0.1 (Figure 4). The predictions for Catchment 3 also turn out to be larger than any of the nearby observed values.

The result can be understood by looking at the nested structure of the catchments in the dataset. Catchment 4 and Catchment 5 cover Catch-

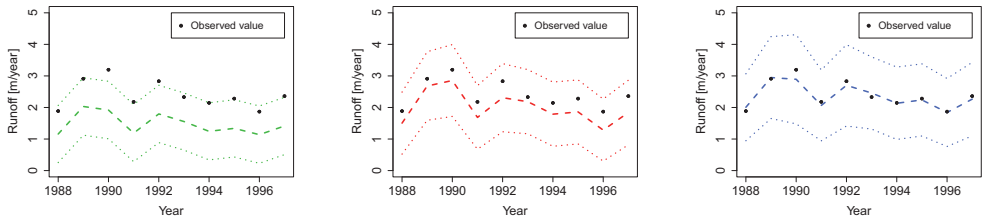


Figure 5: The posterior mean for spatial predictions in Catchment 3 (test **T2**) with corresponding 95 % posterior prediction intervals for observation design P (left), A (middle) and $P + A$ (right).

ment 3, and through our model formulation they put constraints on the total runoff in this area. As Figure 1 shows, there are two precipitation gauges inside Catchment 5 for which the point runoff generated is lower than the mean annual runoff in the surrounding two catchments. To preserve the water balance, the predicted annual runoff in the remaining parts of Catchment 4 and Catchment 5 has to be larger than any of the values that are observed in the surrounding area. This interaction between nested areal observations and point observations makes the model able to correctly identify Catchment 3 as a wetter catchment than any of the nearby catchments, and we have demonstrated that we have a geostatistical model that does more than smoothing.

This does not mean that the interaction between point and areal observations always lead to improved predictions (see e.g. Catchment 5 in Figure 4). However, overall the results in Figure 4 show that on average we benefit from including all available data ($P + A$) in the analysis when making spatial predictions, and that using only point observations gives poor predictions. P performs considerably worse than A, $P + A$ and Top-Kriging for three of the catchments (Catchment 3, 4 and 5)

The scatterplots in Figure 6 compare the spatial predictions from 1988-1997 (**T2**) to the actual observations for each (ungauged) catchment for P, A and $P+A$. Overall, observation designs A and $P+A$ provide predictions that are symmetric around the corresponding observed runoff. However, if we look more closely at the predictions for each catchment, we see that A and $P + A$ tend to either overestimate or underestimate the annual runoff

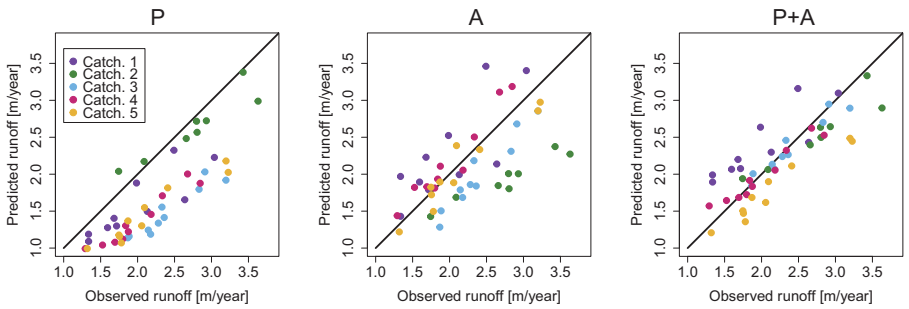


Figure 6: The posterior mean for spatial predictions in ungauged catchments for P (left), A (middle) and $P + A$ (right) compared to the corresponding observed value (**T2**).

within a catchment. This is seen most clearly for Catchment 1 where the annual runoff is overestimated for A and $P+A$, and for Catchment 2 where the runoff is underestimated for A. Top-Kriging is not visualized here, but this reference approach gives similar results as observation scheme A.

The results in Figure 6 show that the same systematic prediction error typically is done each year for a specific catchment. The biases are however small enough that the actual observations are covered by the corresponding 95% posterior prediction intervals for A and $P + A$ for most catchments. This can be seen in Table 2.

For P the situation is different: Figure 6 shows that the annual runoff is underestimated for all catchments. In addition, the posterior standard deviation for runoff is typically unrealistically small for P contributing to narrow posterior prediction intervals. Large biases combined with small posterior standard deviations lead to a low empirical coverage for the spatial predictions for P, and on average the coverage of a 95% posterior prediction interval is as low as 42%. For P, neither the posterior mean nor the posterior variance reflects the properties of the underlying process.

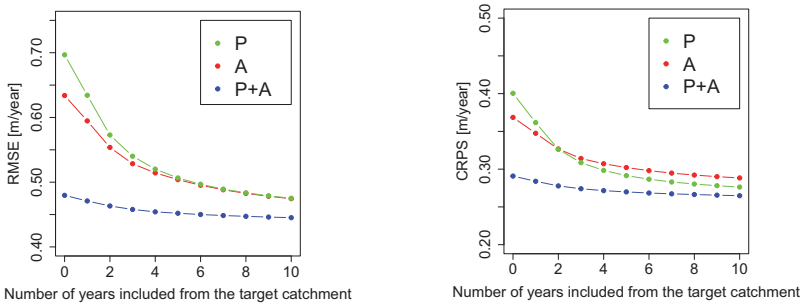
In tests **T3g** and **T3u** annual runoff was predicted for unobserved future years (1998-2014) when 0-10 observations from the target catchment between 1988 and 1997 were included in the likelihood, together with observations of P and/or A from other locations and catchments. The resulting

Table 2: The proportion of the observations that falls into the corresponding 95% posterior prediction interval for spatial predictions of runoff (**T2**) in catchment $\mathcal{A}_1, .., \mathcal{A}_5$ for 1988-1997 when the target catchment is treated as ungauged.

	$\mathcal{A}_1$	$\mathcal{A}_2$	$\mathcal{A}_3$	$\mathcal{A}_4$	$\mathcal{A}_5$	All
P	1	0.5	0.5	0.1	0	0.42
A	1	0.7	1	0.9	1	0.92
$P + A$	1	1	1	1	0.60	0.92

predictive performance is visualized in Figure 7. As for the spatial predictions, $P + A$ gives the lowest RMSE and CRPS on average. For ungauged catchments (when 0 years of observations from the target catchment are included), P and A perform considerably worse than $P + A$. However, when we include some years of observations from the target catchment, we see a large drop in the RMSE and CRPS for P and A. The posterior mean for a future year is given by the posterior mean of $\int_{\boldsymbol{u} \in \mathcal{A}_k} (\beta_c + c(\boldsymbol{u}))d\boldsymbol{u}$, i.e. the plots show that we get a large change in the climatic part of the model when we include information from a new location or catchment.

This result can be understood from the results from the parameter estimation in **T1**: The posterior median of the standard deviation of the climatic GRF σ_c is approximately twice as large as the median of the marginal standard deviation for the annual GRF σ_x for all observation designs (Table 1). Hence, the potential value of a new data point from an unobserved location can be large, as the new observation affects the climatic part of the model that has a substantial impact on the predictions for all years under study. Furthermore, the large spatial climatic effect can also be a possible explanation for the systematic errors we saw for the spatial predictions in Figure 5 and Figure 6 (**T2**). A strong climatic field $c(\boldsymbol{u})$ indicates that the same spatial runoff pattern is repeated each year, and if we fail to characterize it, systematic errors are a reasonable consequence.



(a) RMSE.

(b) CRPS.

Figure 7: Predictive performance for future runoff (1998-2014) in catchment $\mathcal{A}_1, ..., \mathcal{A}_5$ when 0-10 years of observations from the target catchment between 1988 and 1997 are included in the observation likelihood together with other observations of P and/or A (tests **T3u** and **T3g**).

6 Simulation study

One of the objectives of this paper was to show how quantifying long-term spatial variability can be used as a tool for understanding the uncertainty and biases in the modeling of environmental variables. In the case study we have already suggested that a strong climatic field $c(\boldsymbol{u})$ can be an explanation for the systematic over- and underestimation we saw for some of the catchments. In the simulation study we present here, we aim to investigate this further, i.e. we explore if the over- and underestimation actually is a model property, and that it is not only caused by e.g. mismatch between the model and the runoff data in Voss. More specifically, if the true underlying process is driven by two different spatial processes, one climatic (common for all years) and one annual (different each year), can these systematic predictive biases be expected for a given catchment and set of observation locations?

In the simulation study, we explore the model properties for different values of the spatial parameters ρ_c, ρ_x, σ_c and σ_x. The parameters could represent different environmental variables or different study areas. By this, we aim to show what insight one can obtain about a spatio-temporal environ-

mental variable of interest and the corresponding study area by separating climatic spatial variability from year dependent effects.

6.1 Experimental set-up

In the simulation study, we simulate from the model described in Section 4 for 9 different configurations of the range parameters ρ_c, ρ_x and the marginal standard deviations σ_c and σ_x. These are shown in Table 3. We here refer to the proportion $\sigma_c^2/(\sigma_c^2+\sigma_x^2)$ as the *climatic spatial dominance* as it represents a quantification of how large the climatic spatial effect $c(\boldsymbol{u})$ is relative to the year specific spatial effects $x_j(\boldsymbol{u})$. Note that Parameter set 1 with $\sigma_c = 0.8$, $\sigma_x = 0.3$, $\rho_c = 20$ and $\rho_x = 100$ corresponds to the posterior medians obtained for the real case study for $P+A$ (Table 1). The other parameter sets could represent the dependency structure of another climatic variable, e.g. temperature or *monthly* runoff, or the annual runoff in another part of the world.

Table 3: Parameters used for the simulation study. Parameter set 1 corresponds to the parameters obtained for the case study for $P + A$ in Table 1. We refer to the proportion $\sigma_c^2/(\sigma_c^2 + \sigma_x^2)$ as the climatic spatial dominance.

Parameter set	σ_c [m/year]	σ_x [m/year]	ρ_c [km]	ρ_x [km]	$\sigma_c^2/(\sigma_c^2 + \sigma_x^2)$
1	0.8	0.3	20	100	0.88
2	0.5	0.5	20	100	0.50
3	0.3	0.8	20	100	0.12
4	0.8	0.3	50	100	0.88
5	0.5	0.5	50	100	0.50
6	0.3	0.8	50	100	0.12
7	0.8	0.3	100	100	0.88
8	0.3	0.5	100	100	0.50
9	0.5	0.8	100	100	0.12

The remaining two parameters are set to $\beta_c = 2$ and $\tau_\beta = 5$ for all experiments, i.e. similar to the posterior medians for $P + A$ in Table 1. Furthermore, we assume that the measurement errors of the point observations are normally distributed with standard deviation 15% of the corresponding simulated value, while the measurement errors of the areal

observations are normally distributed with standard deviation 3% of the corresponding simulated value. These estimates are set based on recommendations from the data provider NVE regarding the measurement errors we typically see for precipitation and runoff.

For all 9 parameter configurations, annual runoff is simulated for the point and areas in Figure 1. This way we obtain a realistic distribution of observations. In total 50 datasets were generated for each parameter set, i.e. there are 50 simulated climates $c(\boldsymbol{u}) + \beta_c$, and for each climate there are 10 replicates of the year specific component $x_j(\boldsymbol{u}) + \beta_j$.

In our experiments, we predict runoff for two of the catchments in Figure 1: Catchment 1 that is not nested and located relatively far from most point observations, and Catchment 4 that is nested and located in the middle of the study area with many surrounding observations. In turn, Catchment 1 or Catchment 4 is left out of the dataset, and 10 years of annual runoff (1988-1997) are predicted for the target catchment based on all point observations and the remaining areal observations from the same time period (1988-1997). That is, we use the setting $P+A$ for all simulated experiments. Furthermore, the predictions are done both when the target catchment is treated as ungauged with 0 annual runoff observation included in the likelihood, and when the target catchment is treated as partially gauged with 1 randomly drawn annual runoff observation (out of 10 years) included in the likelihood.

In order to investigate the relationship between the model parameters and prediction bias over time, we quantify bias as follows: For each of the 50 climates, we predict runoff for Catchment 1 and Catchment 4 for 10 years. Then, we compute the empirical probability that all of the 10 true (simulated) values of annual runoff are either below or above the 10 corresponding posterior medians for a specific catchment. We refer to this as the probability of *systematic bias*, i.e.

Prob(Systematic bias) = Prob(All 10 simulated values are either below or above the 10 posterior medians).

Systematic biases were common in the case study, and can be seen for example for Catchment 5 in Figure 6 for P+A. We report the probability

of systematic bias as *one* value per parameter set, estimated based on 100 events (50 climates and 2 target catchments).

6.2 Results from the simulation study

We first present the overall 95 % coverage for the simulation study, based on predictions of 10 years of runoff for 2 catchments and 50 climates. These are shown in Table 4, and we find that the empirical coverages are close to 95 % for all the parameter sets in Table 3. If we next consider a scatter plot of the 1000 true and predicted values (not included here), the predictions are also unbiased with respect to the true runoff values. The 95 % coverages and the scatter plots confirm that the model behaves as expected asymptotically for all parameter sets.

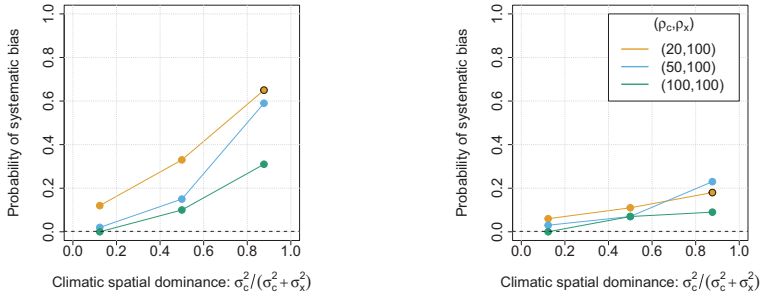
Table 4: Overall 95 % coverage for the simulation study over 50 climates, 2 catchments and 10 years of predictions for $P + A$. For ungauged catchments, there are 0 observations from the target catchment in the likelihood while for partially gauged catchments there is 1 annual observation available from the target catchment.

Parameter set	1	2	3	4	5	6	7	8	9
Ungauged catchments	0.96	0.96	0.94	0.94	0.98	0.96	0.96	0.96	0.96
Partially gauged catchments	0.93	0.93	0.94	0.95	0.95	0.96	0.96	0.96	0.95

Next, Figure 8a shows an visualization of the systematic bias obtained for the simulation study when the target catchments (Catchment 1 and Catchment 4) are treated as ungauged. Recall that systematic bias here is measured as the probability that all 10 true annual runoff values are either below or above the corresponding predicted value for a specific climate and catchment. We see a clear relationship between this bias and the climatic spatial dominance given by the proportion $\sigma_c^2/(\sigma_c^2 + \sigma_x^2)$: When annual spatial effects dominate over climatic spatial effects and $\sigma_c \ll \sigma_x$, the probability of systematic bias is close to zero (around 0.2 %). However, when most of the spatial variability is due to the climate ($\sigma_c \gg \sigma_x$), this probability increases to 30-65% depending on the values of the range parameters ρ_c and ρ_x. For the parameters corresponding to the Norwegian case study, the probability of systematic bias was 65 %. Hence, systematic errors like we saw for e.g. Catchment 5 (P+A) in Figure 6, can be expected

quite often for these parameter values. Figure 8a also shows that the probability of systematic bias is largest when the climatic range ρ_c is low, i.e. when the information gain from the neighboring catchments is low.

From a statistical point of view, the above results are intuitive: If most of the spatial variability can be explained by climatic conditions, there are large dependencies between years. Either we typically perform accurate predictions all years, or poor predictions all years. Considering all ungauged catchments in Norway, we can expect that 95% of the true runoff values are inside the corresponding 95% posterior prediction intervals on average (Table 4), but if we consider predictions for individual catchments over time, a large proportion of the predictions will be biased in one direction or the other (Figure 8a). The simulation study shows that the systematic bias we obtained for the case study are not necessarily a result of mismatch between the data and the fitted model, but can indeed be a result of the strong climate around Voss ($\sigma_c \gg \sigma_x$).



(a) 0 observations from the target catchment.

(b) 1 observation from the target catchment.

Figure 8: The empirical probability that all of the 10 true annual values are either above or below the posterior median value for Catchment 1 and Catchment 4 over 50 climates for ungauged catchments (Figure 8a) and partially gauged catchments (Figure 8b). The black circle corresponds to the parameter values we have for the case study from Voss. The black dashed line is the theoretical probability that all the observed values are above or below the posterior median when studying a process that actually is independent over years ($2 \cdot 0.5^{10} = 0.2\%$). This is included as a reference.

So far we have considered the probability of systematic bias when there are no data from the target catchments available. Next, in Figure 8b, we present the probability of systematic bias when there is one annual observation included in the likelihood. For $\sigma_c \ll \sigma_x$, i.e. when $\sigma_c^2/(\sigma_c^2 + \sigma_x^2)$ is close to zero, we see that the probability of systematic bias in general is low for both ungauged catchments (Figure 8a) and partially gauged catchments (Figure 8b). For this scenario, a new data point from the target catchment don't have a considerable impact on the probability of systematic bias. However, if $\sigma_c \gg \sigma_x$ as in Voss, we find that the extra data point on average leads to a large reduction in the systematic bias probability in Figure 8b compared to the systematic bias probability we saw for the ungauged catchments in Figure 8a. This is found for all combinations of ρ_c and ρ_x, but the tendency is strongest if $\rho_c \ll \rho_x$ as in Voss. The results in Figure 8 are thus comparable to the results in Figure 7 for the case study, and illustrate the potential value of data from a new location for different parameter values.

7 Discussion

In this paper we have presented a model for annual runoff that consistently combines data of different spatial support. The suggested model is a geostatistical model with two spatial effects: A climatic long-term effect and a year dependent effect that describes the annual discrepancy from the climate. The model was used to estimate mean annual runoff in the Voss area in Norway.

The main focus of the study was on exploring how the combination of point and nested areal observation affects runoff predictions, to demonstrate that our model has mass-conserving properties and to show how quantifying long-term spatial variability can be used as a tool for understanding biases in environmental modeling and for exploiting short records of data. There are three key findings: 1) On average we benefit from including all available observations in the likelihood, both point and areal data. $P+A$ performed better than P and A in terms of RMSE, CRPS and the coverage of the 95 % posterior prediction intervals in our case study. $P+A$ also performed better than the referenced method Top-Kriging that

only supports areal observations. 2) The suggested model that combines point and areal observations is particularly suitable for modeling the nested structure of catchments. The case study showed that the model was able to identify Catchment 3 as a wetter catchment than any of the surrounding catchments and precipitation stations. This was a consequence of using information from two overlapping catchments to constrain and distribute the annual runoff correctly. The interaction between the point and nested areal observations gives a geostatistical model that does more than smoothing. The linear constraints also represent a main difference to Top-Kriging that does not ensure mass preserved predictions. 3) How dominating climatic spatial effects are compared to annual spatial effects has a large influence on the predictability of runoff. If most of the spatial variability can be explained by long-term (climatic) weather patterns and processes, systematic biases for a location over time can be expected as long as the same observation design is used.

The fact that $P + A$ performed better than A for most catchments around Voss, indicates that the point and areal observations of runoff were sufficiently compatible for most catchments, i.e. that evaporation subtracted from precipitation was a valid approximation of point runoff. This interpretation of point runoff is reasonable in areas like Voss where the annual precipitation is considerably larger than the annual evaporation. The evaporation data are uncertain and should not make a large impact on the resulting predictions. In many areas of the world, the observed annual evaporation is more than 50 % of the annual observed precipitation. In such areas, our framework could provide negative point observations and results that are hard to interpret. Negative runoff can in general be a problem in our Gaussian model. Log transforming the data is a solution if considering only point data (P), but is not an option when modeling areal data (A and $P + A$) because the log transformation does not work well with the linear aggregation in Equation (9). For areas with observed values close to zero, extra caution should therefore be taken regarding negative, non-physical results. To avoid negative predictions it is also important to make sure that the mesh used in the SPDE approach (Figure 2b) is fine enough to capture the rapid spatial variability in the study area.

Precipitation observations are often avoided as an information source when

performing interpolation of runoff in hydrological applications, but the results presented here show that the point observations can contain valuable information when used together with areal observations. At least in data sparse areas with few streamflow observations. However, there is still room for improvement in the compatibility between the two observation types: The observation designs including only point observations P provided a clear underestimation of annual runoff for most catchments in the case study. It was also seen that the spatial field provided by the precipitation observations (P) was smoother than the spatial field provided by the runoff observations (A) in Figure 3a, which is a typical result: The increase in spatial variability from precipitation to runoff is mainly explained by small scale variability introduced by soil and vegetation (Skøien et al., 2003b). Consequently, if the point data are allowed to dominate over the areal data, the point data can cause a runoff field that is too smooth, which affects both the posterior mean and the posterior standard deviation disadvantageously.

Furthermore, it is worth mentioning that all of the available precipitation gauges are located at a lower elevation than the mean elevation of the five catchments in the dataset. This is a common problem. Precipitation gauges are often located at low elevations, close to settlements where the gauges are easy to maintain. It is known that the amount of precipitation typically increases with elevation. There is therefore a lack of information about precipitation at high elevations in the data. Adding the fact that the precipitation gauges often fail to catch a large proportion of the precipitation, in particular when it comes as snow and it is windy (Kochendorfer et al., 2017), essential information about the precipitation and runoff field could be lost. To solve the compatibility issues, elevation was considered as a covariate in a preliminary study (Ødegård, 2017), but this did not lead to significant improvements, and the results are not included here. Another option could be a preferential sampling approach where we assume that the locations of the precipitation gauges are distributed according to a log-Gaussian Cox process that depends on the response variable, here through elevation implicitly (Diggle et al., 2010).

Elevation is also known to be a factor that affects the spatial dependency structure of precipitation, and Voss is a mountainous area. The spatial

range is typically larger in lowlands and decreases with elevation. A non-stationary model similar to the one presented in Ingebrigtsen et al. (2014) with a range and a marginal variance that changes with elevation could be considered. This can easily be implemented within the INLA-SPDE framework. However, in this case the dataset is small and the complexity of the spatial variability large. We also have a model with only one replication of the climatic spatial effect which was the dominating spatial component. A non-stationary model would probably be too complicated and lead to identifiability issues (Ingebrigtsen et al., 2015).

Regardless of the increased complexity in an extended model, it is reasonable to believe that an accurate representation of the climatic conditions at a target location is crucial when predicting annual runoff and other climate related variables. In the simulation study, we demonstrated how systematic under- and overestimation of a target variable can be expected over time when we fail to characterize the underlying climate in areas where the climatic spatial field's marginal standard deviation σ_c is large relatively to the other model standard deviations. We also found a clear relationship between the model parameters of the suggested model, and systematic prediction bias over time. This shows that the two field model (and its parameters) can contribute with useful insight about the properties of a study area and/or an environmental variable of interest.

In spite of the large biases documented for annual runoff predictions in this article, a dominating climate also gives opportunities. In this article a model with a climatic component was suggested. The climatic component included a spatial effect that was common for all years of observations. This component made it relatively simple to exploit short records of data, and the runoff predictions could easily be improved by including a few observations from the target catchments. Time series from several years are not needed because one or two observations from a new catchment updates the climatic component that has a large impact to the final model if σ_c dominates over the other model variances. Here, we again note how the model parameters can contribute with useful information about the study area and/or the environmental variable of interest: The potential gain of collecting a new data point from a new location, i.e. a short record, can be indicated from the spatial parameters, in particular from the proportion

$\sigma_c^2/(\sigma_c^2 + \sigma_x^2)$.

The ability to exploit short records is another main benefit of the suggested model over existing spatial models used for runoff interpolation, like e.g. Top-Kriging. For practitioners, a model with the described properties can be useful in situations where there exist one or few observations from a catchment of interest. Short duration runoff observations are quite common in hydrological datasets, e.g. from planned short duration missions for water resources assessments, or from gauging stations that are closed after a revision of the gauging network. Large infrastructure projects measuring a few years of annual runoff for a relevant catchment is also achievable.

References

K. Adamowski and C. Bocci. Geostatistical regional trend detection in river flow data. *Hydrological Processes*, 15(18):3331–3341, 2001. doi: 10.1002/hyp.1045.

S. Beldring, K. Engeland, L. A. Roald, N. R. Sælthun, and A. Voksø. Estimation of parameters in a distributed precipitation-runoff model for norway. *Hydrology and Earth System Sciences*, 7(3):304–316, 2003.

M. Blangiardo and M. Cameletti. *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Wiley, 1st edition, 2015.

G. Blöschl, M. Sivapalan, T. Wagener, A. Viglione, and H. Savenije. *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales*. Camebridge University press, 2013.

S. Brenner and L. Scott. *The Mathematical Theory of Finite Element Methods, 3rd Edition. Vol. 15 of Texts in Applied Mathematics*. Springer, 2008.

N. Cressie. *Statistics for spatial data*. J. Wiley & Sons, 1993.

P. J. Diggle, R. Menezes, and Ting-li Su. Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59 (2):191–232, 2010. doi: 10.1111/j.1467-9876.2009.00701.x.

M.B. Fiering. Use of correlation to improve estimates of the mean and variance. *USGS Publications Warehouse*, 1963.

G-A. Fuglstad, D. Simpson, F. Lindgren, and H. Rue. Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association*, 114(525):445–452, 2019. doi: 10.1080/01621459.2017.1415907.

D. Gamerman and H. F Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC, 2006.

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC, 2nd ed. edition, 2004.

T. Gneiting and A. E. Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102:359–378, 2007. doi: 10.1198/016214506000001437.

L. Gottschalk. Interpolation of runoff applying objective methods. *Stochastic Hydrology and Hydraulics*, 7:269–281, 1993.

P.Y. Groisman and D. R. Legates. The Accuracy of United States Precipitation Data. *Bulletin of the American Meteorological Society*, 75:215–227, 1994.

P. Guttorp and T. Gneiting. Studies in the history of probability and statistics XLIX on the matérn correlation family. *Biometrika*, 93(4):989–995, 2006. doi: 10.1093/biomet/93.4.989.

R. Ingebrigtsen, F. Lindgren, and I. Steinsland. Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, 8:20 – 38, 2014. doi: 10.1016/j.spasta.2013.06.002.

R. Ingebrigtsen, F. Lindgren, I. Steinsland, and S. Martino. Estimation of a non-stationary model for annual precipitation in southern Norway using replicates of the spatial field. *Spatial Statistics*, 14:338 – 364, 2015. doi: 10.1016/j.spasta.2015.07.003.

J. Kochendorfer, R. Rasmussen, M. Wolff, B. Baker, M. E. Hall, T. Meyers, S. Landolt, A. Jachcik, K. Isaksen, R. Brækkan, and R. Leeper. The quantification and correction of wind-induced precipitation measurement errors. *Hydrology and Earth System Sciences*, 21(4):1973–1989, 2017.

E. Krainski, V. GÃ³mez Rubio, H. Bakka, A. Lenzi, D. Castro-Camilo, D. Simpson, F. Lindgren, and H. Rue. *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. 09 2018. doi: 10.1201/9780429031892.

G. Laaha and G. Blöschl. Low flow estimates from short stream flow records — a comparison of methods. *Journal of Hydrology*, 306(1):264 – 286, 2005. doi: 10.1016/j.jhydrol.2004.09.012.

F. Lindgren, H. Rue, and J. Lindström. 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)*, 73:423–498, 2011. doi: 10.1111/j.1467-9868.2011.00777.x.

R. Merz and G. Blöschl. Flood frequency regionalisation - spatial proximity vs. catchment attributes. *Journal of Hydrology*, 302:283 – 306, 2005. doi: 10.1016/j.jhydrol.2004.07.018.

P. Moraga, S. M. Cramb, K. L. Mengersen, and M. Pagano. A geostatistical model for combined analysis of point-level and area-level data using INLA and SPDE. *Spatial Statistics*, 21:27 – 41, 2017. doi: 10.1016/j.spasta.2017.04.006.

Q. Mu, F.A Heinsch, M. Zhao, and S.W Running. *Remote Sensing of Environment*, 111:519–536.

E. L. Neff. How much rain does a rain gage gage? *Journal of Hydrology*, 35:213–220, November 1977.

T. Reitan and A. Petersen-Øverleir. Bayesian methods for estimating multi-segment discharge rating curves. *Stochastic Environmental Research and Risk Assessment*, 23:627–642, 2009.

H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 2005.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71:319–392, 2009. doi: 10.1111/j.1467-9868.2008.00700.x.

E. Sauquet, L. Gottschalk, and E. Lebois. Mapping average annual runoff: A hierarchical approach applying a stochastic interpolation scheme. *Hydrological Sciences Journal*, 45(6):799–815, 2000. doi: 10.1080/02626660009492385.

D. Simpson, H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, 32:1–28, 2017. doi: 10.1214/16-STS576.

J. O. Skøien, R. Merz, and G. Blöschl. Top-kriging - geostatistics on stream networks. *Hydrology and Earth System Sciences Discussions*, 10:277–287, 2006. doi: 10.5194/hess-10-277-2006.

J. O. Skøien, G. Blöschl, and A. W. Western. Characteristic space scales and timescales in hydrology. *Water Resources Research*, 39, 2003a. doi: 10.1029/2002WR001736.

J. O. Skøien, G. Blöschl, and A. W. Western. Characteristic space scales and timescales in hydrology. *Water Resources Research*, 39(10), 2003b.

A. Stohl, C. Forster, and H. Sodemann. Remote sources of water vapor forming precipitation on the Norwegian west coast at 60 °N - a tale of hurricanes and an atmospheric river. *Journal of Geophysical Research: Atmospheres*, 113(D5), 2008. doi: 10.1029/2007JD009006.

A. Viglione, J. Parajka, M. Rogger, J L. Salinas, G. Laaha, M. Sivapalan, and G Blöschl. Comparative assessment of predictions in ungauged basins - Part 3: Runoff signatures

in Austria. *Hydrology and Earth System Sciences Discussions*, 10:449–485, 2013. doi: 10.5194/hessd-10-449-2013.

C. Wang, M.A. Puhan, and R. Furrer. Generalized spatial fusion model framework for joint analysis of point and areal data. *Spatial Statistics*, 23:72 – 90, 2018.

P. Whittle. On stationary processes in the plane. *Biometrika*, 41:434–49, 1954.

P. Whittle. Stochastic processes in several dimensions. *Bulletin of the International Statistical Institute*, 40:974–994, 1963.

E. Widén-Nilsson, S. Halldin, and Chong yu Xu. Global water-balance modelling with wasmod-m: Parameter estimation and regionalisation. *Journal of Hydrology*, 340(1): 105 – 118, 2007. doi: https://doi.org/10.1016/j.jhydrol.2007.04.002.

WMO. *International meteorological vocabulary*. 1992.

M.A Wolff, A. Petersen-Øverleir, K. Ødemark, T. Reitan, and R. Brækkan. Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: Results of a Norwegian field study. *Hydrology and Earth System Sciences*, 19, 2015.

K. Zhang, J. S. Kimball, Q. Mu, L. A. Jones, S. J. Goetz, and S. W. Running. Satellite based analysis of northern ET trends and associated changes in the regional water balance from 1983 to 2005. *Journal of Hydrology*, 379:92 – 110, 2009.

K. Zhang, J. S. Kimball, R. R. Nemani, and S. W. Running. A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006. *Water Resources Research*, 46(9), 2010.

J. Ødegård. Spatial non-stationary models for precipitation with elevation in the dependency structure - a case study of annual precipitation in hordaland. *Master thesis, Norwegian University of Science and Technology*, 2017.

Paper B

Estimation of annual runoff by exploiting long-term spatial patterns and short records within a geostatistical framework

Thea Roksvåg, Ingelin Steinsland and Kolbjørn Engeland

In revision in a hydrology journal

Estimation of annual runoff by exploiting long-term spatial patterns and short records within a geostatistical framework

Thea Roksvåg[1], Ingelin Steinsland[1] and Kolbjørn Engeland[2]
[1] Department of Mathematical Sciences, NTNU, Norway
[2]The Norwegian Water Resources and Energy Directorate (NVE)

Abstract

In this article, we present a Bayesian geostatistical framework that is particularly suitable for interpolation of hydrological data when the available dataset is sparse and includes both long and short records of runoff. A key feature of the proposed framework is that several years of runoff are modeled simultaneously with two spatial fields: One that is common for all years under study that represents the runoff generation due to long-term (climatic) conditions, and one that is year specific. The climatic spatial field captures how short records of runoff from partially gauged catchments vary relative to longer time series from other catchments, and transfers this information across years. To make the Bayesian model computationally feasible and fast, we use integrated nested Laplace approximations (INLA) and the stochastic partial differential equation (SPDE) approach to spatial modeling.

The geostatistical framework is demonstrated by filling in missing values of annual runoff and by predicting mean annual runoff for around 200 catchments in Norway. The predictive performance is compared to Top-Kriging (interpolation method) and simple linear regression (record augmentation method). The results show that if the runoff is driven by processes that are repeated over time (e.g. orographic precipitation patterns), the value of including short records in the suggested model is large. For partially gauged catchments the suggested framework perform better than comparable methods, and one annual observation from the target catchment can lead to a 50 %

reduction in RMSE compared to when no observations are available from the target catchment. We also find that short records safely can be included in the framework regardless of the spatial characteristics of the underlying climate, and down to record lengths of one year.

1 Introduction

Characteristic values for streamflow are used for various purposes in water resources management. High flow indices or design flood estimates are needed for flood risk assessments and design of infrastructure and dams, low flow indices are needed for assessment of environmental flow and reliability assessment of water supply, while mean annual flow is an important basis for water resources management and a key for design of water supply systems and allocation of water resources between stakeholders. Mean annual flow can also be used as a predictor for low flow and high flow indices (Sælthun et al., 1997; Engeland and Hisdal, 2009).

At locations with measurements, the streamflow indices can be estimated based on observations. However, streamflow is only measured at a limited number of locations, and in many applications we need to predict the streamflow indices at ungauged locations. This is a central problem in hydrology and known as the Prediction in Ungauged Basins problem (Blöschl et al., 2013). Often it is of interest to estimate flow indices that represent the long-term average behavior in a catchment. If this is the case, using only a few years of data from the target catchment might lead to biased estimates. The reason is climate variability over short time scales combined with sample uncertainty. Often a minimum record length is recommended for estimation of long-term indices, but a substantial part of the available streamflow gauges in the world have too short records to provide reliable estimates. These short data series can, however, provide useful information if they are used together with longer time series from other catchments (Laaha and Blöschl, 2005). Motivated by this, we propose a framework for runoff interpolation particularly suitable for datasets including data series of this type, more specifically runoff datasets including a mix of fully gauged catchments (with data available from the whole study period) and partially gauged catchments (with data available from

a subset of the study period). We suggest a framework for runoff interpolation that unifies two commonly used statistical approaches for runoff estimation: Geostatistical approaches and approaches for exploiting short records of data.

Within the geostatistical framework, Gaussian random fields (GRFs) are often used to model hydrological phenomena that are continuous in space and/or time. The hydrological variable of interest is a GRF if a vector containing a random sample of length n from the process follows a Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ (Cressie, 1993). The elements in the covariance matrix are typically determined by a covariance function that is specified based on the pairwise distances between the n target locations. For most environmental variables it is straight forward to compute these distances. However, for runoff related variables the measure of distance is ambiguous because the observations are related to catchment areas, some of them nested, and not to point locations in space. Traditionally, this challenge has been solved by simply interpreting runoff as a point referenced process linked to the catchment centroids or stream outlets (see e.g. Merz and Blöschl (2005); Skøien et al. (2003); Adamowski and Bocci (2001)). The problem with these methods is that they can lead to a violation of basic conservation laws, and several alternatives approaches are suggested for making an interpolation scheme that takes the nested structure of catchments into account (Sauquet et al., 2000; Gottschalk, 1993; Skøien et al., 2006). In particular, the Top-Kriging approach suggested by Skøien et al. (2006) has shown promising results for interpolation of hydrological variables (Viglione et al., 2013). In the Top-Kriging approach, information from a subcatchment is weighted more than information from a nearby non-overlapping catchment when performing runoff predictions for an ungauged catchment.

In the literature, there exist several techniques to exploit short records of runoff, and these are known as record augmentation techniques. The first step in a record augmentation procedure is often to find one or several donor catchments with longer time series of runoff. The donor catchments are typically selected based on runoff correlation, catchment similarity, or proximity in space. By applying e.g. linear regression approaches and/or computing the correlation between time series, a relationship between the

target catchment and the donor catchments is developed. Next, the longer time series from the donor catchment(s) are used to perform predictions for the target catchments for years/months/days without measurements (see e.g. Fiering (1963), Hirsch (1982), Matalas and Jacobs (1964), Vogel and Stedinger (1985) or Laaha and Blöschl (2005)). The regression and/or correlation analysis is performed based on runoff observations that is of the same type as the target flow index, i.e. for annual runoff, short records of annual runoff are used (Blöschl et al., 2013).

In this paper, we suggest a geostatistical Bayesian framework that represents a new way of exploiting short records of data. The framework is constructed to exploit long-term spatial patterns stored in sparse datasets, i.e. hydrological datasets with several missing values. A key feature of the suggested framework is that it simultaneously models several years of runoff. This is done by using two statistical spatial components or GRFs in the hydrological model: The first GRF is common for all years under study and models the long-term spatial variability of runoff. We denote this the climatic GRF as it represents the spatial variability over time, or what we refer to as the climate in the study area. In this context the term *climate* also includes the runoff generation due to catchment characteristics that are static, like elevation and slope. The other GRF is year-specific and models the annual discrepancy from the climate, and we denote this the annual or year-specific GRF. If we have a study area for which the spatial variability of runoff is stable over time, the climatic GRF will capture this tendency. Hence, it will also capture how short records of runoff vary relative to longer data series from other catchments. On the other hand, if there are no strong long-term trends present in the data, the year-specific GRF will dominate over the climatic GRF. For this scenario, short records from the target catchment(s) will have less impact on the final results. By adjusting the two spatial fields relative to each other, our method represents a way for detecting long-term trends and uses this to exploit short records in the runoff interpolation.

The framework we suggest is flexible and can be used for any hydrological variable. However, its benefits are linked to exploiting long-term spatial trends in the data, and in order to work better than other interpolation methods, the hydrological variable of interest should be driven by processes

that are repeated over time. For this reason, we develop our methodology for *annual* runoff. This is a flow index that often has a prominent spatial pattern over years, for example due to orographic precipitation and topography that creates weather divides. To describe study areas and/or variables like this, we hereby introduce the terms *hydrologically spatially stable* and *hydrological spatial stability*. For hydrologically spatially stable areas, the difference in runoff between two locations for a given year is close to the difference in runoff between these two locations any other year. Be aware that a hydrologically spatially stable area can both have large differences in annual runoff between two close locations, and have large variability in annual runoff over years for a given location. The key property is that the underlying spatial pattern is preserved over time.

While annual runoff represents a hydrologically spatially stable variable for many countries, the spatial pattern for monthly runoff is typically less stable. This is due to local weather patterns and the variability in the seasonality of snow accumulation and snow melt. To demonstrate our methodology for a variable with less hydrological spatial stability, we therefore fit the framework to annual time series of monthly runoff. These predictions allow us to discuss how the approach might work in different regions.

In the following presentation, we introduce two versions of our framework, i.e. two geostatistical models. The first model we propose is denoted the areal model and is particularly suitable for mass-conserved hydrological variables. It ensures that the water balance is preserved for the predicted runoff for any point in the landscape, and defines the average runoff in a catchment as the average point runoff integrated over (nested) catchment areas. This way, the nested structure of catchments is taken into account, and the interpretation of covariance between two catchments is similar to the one of Top-Kriging. The areal model for annual runoff is already presented in Roksvåg et al. (2020) where its mass-conserving properties were demonstrated through an example from Voss in western Norway. The model's ability to exploit short data records was also indicated in Roksvåg et al. (2020), but the property was not tested for a larger dataset or compared to any existing methods. This is a key contribution of this article.

As an alternative to the areal model, we also propose a model that de-

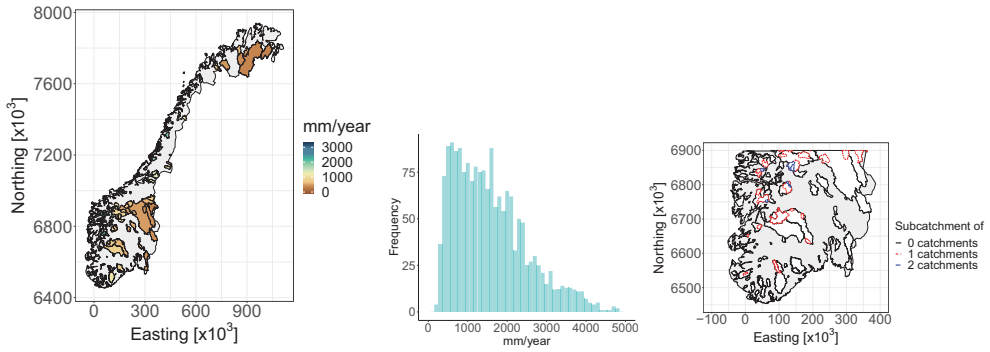
fines runoff as a point referenced process for which distances are measured between the catchment centroids. This model does not consider preservation of water balance, but on the other hand it can be used for any point referenced environmental variable, and it is computationally faster than the areal model. This model is more similar to models that have been used traditionally in hydrology, and we denote this the centroid model. Both the areal model and the centroid model have the ability to exploit hydrological spatial stability, but have different benefits, drawbacks and hence also area of use. These are discussed and highlighted throughout the article.

The main objective of this work is to present and evaluate the new geostatistical framework for exploiting short records and to compare its performance to Top-Kriging (interpolation method) and simple linear regression (record augmentation technique). In particular our goals are to:
1) Assess the two spatial models' ability to fill in missing annual observations of runoff for ungauged and partially gauged catchments.
2) Assess the two spatial models' ability to predict *mean* annual runoff for a longer time period for catchments with varying record lengths.
Through 1) and 2) we also aim to:
3) Demonstrate the potential added value of including short records in the modeling, compared to not using them or compared to using traditional methods.

The framework is evaluated by using annual and monthly runoff data from catchments in Norway. This dataset is presented in the section that follows (Section 2). Next, in Section 3, we briefly introduce relevant statistical background theory and notation. In Section 4 the suggested model for annual runoff is presented, before evaluation scores and experimental set-up are presented in Section 5. Here, we have one experimental set-up for annual predictions (Section 5.1) and one set-up for mean annual predictions (Section 5.2). In Section 6, the results are presented before they are discussed in Section 7. Finally, we conclude in Section 8.

2 Study area

The study is carried out by using a dataset from Norway provided by the Norwegian Water Resources and Energy Directorate (NVE). It originally consisted of *daily* runoff data from 1981-2010. To make the data suitable for an analysis, a data preparation procedure was performed to construct datasets for two purposes: For assessing the framework's ability to fill in missing annual data and for assessing the framework's ability to predict mean annual runoff.



(a) Mean annual runoff for 1996-2005.

(b) Annual runoff from 1995-2006.

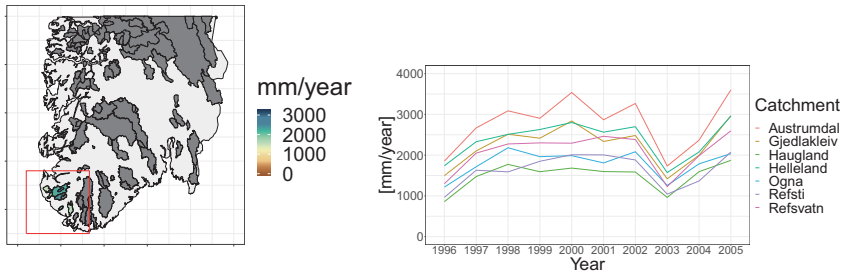
(c) Nested catchments.

Figure 1: Mean annual runoff (1996-2005) from 180 fully gauged catchments in Norway (1a) and annual runoff observations from all 180 catchments and years (1b). These data are used to evaluate the framework's ability to fill in missing values for individual years. 30 % of the involved catchments are nested and most of these are located in southern Norway as visualized in Figure 1c. In this figure, colored catchments are subcatchments of at least one larger catchment, while the black catchments are not subcatchments of any larger catchment (but might contain 1 or 2 smaller catchments). In the visualization in Figure 1a, subcatchments are plotted on top of larger catchments, and this is done throughout the article. The coordinate system used is EUREF89 - UTM33N (EPSG 25833). See Figure 7 for a closer image of the observed mean annual runoff in southern Norway (1996-2005).

To make a cross-validation dataset for the experiments related to infill of missing annual data, the daily runoff data were aggregated to annual runoff

data for hydrological years that start September 1th and end August 31st. We chose to consider a study period from 1996-2005: For this period we had the maximum number of fully gauged catchments, i.e. 180 catchments. These 180 fully gauged catchments have areas ranging from 13 km^2 to 15500 km^2 and median elevations from 85 to 1562 m a.s.l. Among these, none were significantly influenced by human activities in the time period of interest. Regulated catchments were removed from the original dataset.

Figure 1a and Figure 1b show two visualizations of the annual data from the 180 Norwegian target catchments. We see a large spatial variability of runoff. The annual runoff (for individual years) ranges from 170 mm/year to 5050 mm/year, whereas the mean annual runoff ranges from 350 mm/year to 4230 mm/year, with the highest values of runoff in western Norway and more moderate values in east and north. In total 53 of the 180 catchments were nested with at least one other catchment, i.e. the degree of nestedness is 30 %. Most of these are located in southern Norway, and the nested structure here is shown in Figure 1c. The remaining 127 catchments did not overlap with any other catchment.



(a) 7 catchments.

(b) Annual time series.

Figure 2: Time series of annual runoff from 7 selected catchments in western Norway. The 7 lines are almost parallel (and barely cross) indicating that most of the spatial variability can be explained by long-term spatial patterns. This represents a good example of what we mean by hydrological spatial stability.

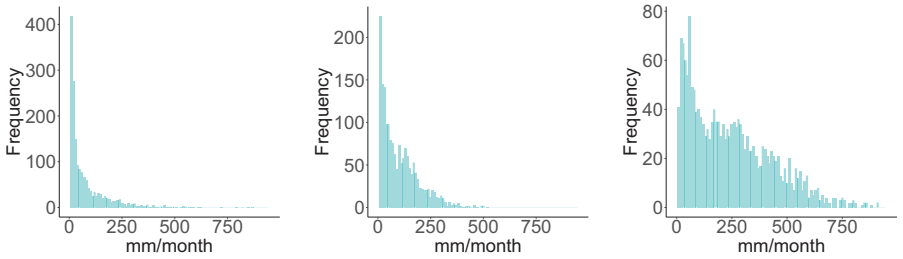
In the Norwegian annual data in Figure 1a we see an east-west pattern of runoff. This is mainly caused by orographic enhancement of frontal precipitation formed around extratropical cyclones. The orographic enhancement is driven by the steep mountains in western Norway that create a

topographic barrier for the western wind belt, which transports moist air across the North Atlantic (Stohl et al., 2008). Due to the orographic enhancement, the maximum precipitation is observed at distances 30-70 km from the coast (Førland, 1979) and not necessarily at the highest elevations since the air dries out due to precipitation. The topography results in a spatial pattern of runoff that is stable over years, which means that there exist long-term spatial patterns in the data that can be exploited.

Figure 2 shows time series of annual runoff from seven catchments in the south-western part of the country. We see a year to year variation for all catchments that is quite large. However, the seven time series are almost parallel (and almost never cross), indicating that the difference in annual runoff between stations is approximately constant over time. Hence, this is a good example of what we mean by hydrological spatial stability. The tendency we see in Figure 2 is typical for the annual runoff in many of the areas in Norway.

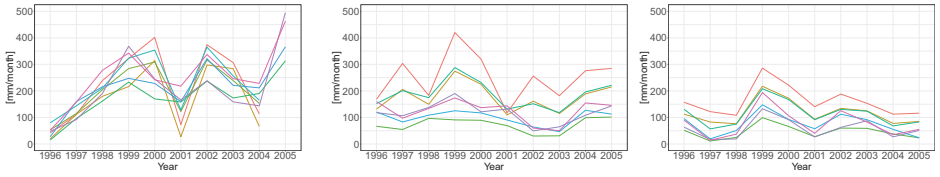
To illustrate the framework's properties for study areas and/or variables that are driven by more unstable weather patterns or hydrological processes, we also aggregated the daily runoff data to monthly runoff for the 180 catchments in Figure 1a. From this we made annual time series of monthly runoff for 1996-2005 for three months: A winter month dominated by snow accumulation (January), a spring month with snow melting (April) and a summer month dominated by rain (June). The annual observations of monthly runoff for the selected months are presented in Figure 3, and we see that January has the lowest average runoff whereas June has the highest. The variation in average monthly runoff describes a runoff regime, and in Norway the combination of snow accumulation, snow melt, and evapotranspiration processes control this regime (Gottschalk et al., 1979). Along the west coast, the winter weather is typically rainy with temperatures above the freezing point. In these regions the highest monthly runoff is observed in October - December. The colder areas are found in the interior of the country with winters dominated by snow accumulation. In these regions the highest monthly runoff is observed for the snow melt season (May – June).

Annual time series of monthly runoff from the 7 selected catchments from Figure 2a are shown in Figure 4. We see that the spatial pattern is less



(a) January. (b) April (c) June

Figure 3: Monthly runoff data (1996-2005) from 180 catchments in Norway for January, April and June. These are used to evaluate the framework's ability to fill in missing values for hydrological variables and/or study areas that are driven by more unstable weather patterns.



(a) January. (b) April. (c) June

Figure 4: Annual series of monthly runoff for January, April and June for the 7 catchments in Figure 2a. The time series for January and April are less parallel compared to the time series for June and for the annual runoff (Figure 2b). This suggests that the datasets from January and April represent a more hydrologically spatially unstable setting.

stable on a monthly scale compared to the annual scale, particularly for January: The difference in monthly runoff between stations over time is not approximately constant for January, and the runoff in January hence represents a more hydrologically spatially unstable variable in Norway. For June however, the hydrological spatial stability is higher.

The cross-validation datasets described so far are used to assess the framework's ability to fill in missing annual observations for a 10 year period and to illustrate how the models behave for different hydrological settings. In addition, we also evaluate the framework's ability to predict *mean* annual

runoff, which is a key hydrological signature. This is done for a 30 year period, from 1981 to 2010. As we consider a longer time period for this assessment, a different subset of the original dataset was used: More specifically annual data from 260 catchments located in southern Norway. These are shown in Figure 5. Each of the 260 catchments in Figure 5 have at least one observation of annual runoff between 1981-2010, but only 83 of them are fully gauged in the time period of interest (i.e. have annual observations for all 30 years). Among the partially gauged catchments, the mean record length is 15, while the median record length is 13. Furthermore, 20 of the involved catchments only have 1, 2 or 3 annual observations. As for the previously described datasets, we removed regulated catchments that were significantly influenced by human activity. Also note that we in this experiment only consider catchments from southern Norway. This is done to reduce the computational complexity of fitting 30 years of runoff simultaneously in a cross-validation setting.

When using the data in Figure 5 to predict mean annual runoff, we do predictions by cross-validation for the 83 fully gauged catchments. However, data from both partially gauged and fully gauged catchments are included in the observation sample (see Section 5.2). For the 83 fully gauged catchments in Figure 5, 53 % of the catchments were nested with a fully gauged or a partially gauged catchment.

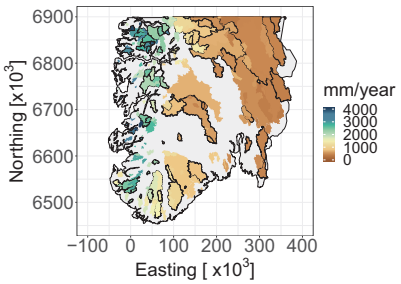


Figure 5: Mean annual runoff for 1981-2010 for 260 catchments in southern Norway where only 83 of them are fully gauged (i.e. have annual data for each year in the study period). The 83 fully gauged catchments have black borders in the above plot. In addition, there are data available from 177 so-called partially gauged catchments. These have at least one annual observation between 1981-2010 and are visible as catchments without borders in the above figure. Among the 83 fully gauged catchments, 44 catchments are nested (53 %) while 39 catchments don't overlap with any other catchment in the dataset. Data from the catchments in this figure are used to evaluate the framework's ability to estimate *mean annual runoff*.

3 Statistical methodology

In Section 4 we present two Bayesian geostatistical models for runoff interpolation particularly suitable for sparse datasets containing several missing values. First, some statistical background is necessary.

3.1 Bayesian statistics and hierarchal modeling

The goal in hydrology is to learn about processes related to hydrological variables like daily rainfall, annual runoff or the 5th percentile flow. To gain knowledge about the different hydrological processes, relevant data are collected. There are always uncertainties related to the data that must be accounted for in an analysis, and which make a statistical analysis appropriate.

Assuming $\boldsymbol{x} = (x_1, ..., x_n)$ is a vector consisting of hydrological variables of interest, e.g. the annual runoff at several locations for a specific year, the

observation likelihood $\pi(\boldsymbol{y}|\boldsymbol{x})$ expresses how the data $\boldsymbol{y} = (y_1, ..., y_n)$ are connected to the truth $\boldsymbol{x}$. In the classical frequentist stastistical approach, the variables in $\boldsymbol{x}$ are considered as unknown, but fixed. In the Bayesian approach however, $\boldsymbol{x}$ is considered to be a quantity whose variation can be described by a probability distribution (see e.g. Casella and Berger (1990)). Prior to the analysis, this probability distribution is expressed through what is called a prior distribution $\pi(\boldsymbol{x})$. This is constructed based on expert knowledge about the variable(s) of interest. The goal of the Bayesian analysis is to update the prior distribution by using data. Through Bayes' formula, the so-called posterior distribution of $\boldsymbol{x}$ is obtained:

$$\pi(\boldsymbol{x}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{x})\pi(\boldsymbol{y}|\boldsymbol{x})}{\pi(\boldsymbol{y})} \propto \pi(\boldsymbol{x})\pi(\boldsymbol{y}|\boldsymbol{x}). \tag{1}$$

Next, the marginal distribution $\pi(x_i|\boldsymbol{y})$ for $x_i \in \boldsymbol{x}$ can be integrated out, and a prediction of x_i can be summarized through e.g. the mean, median or the mode of the posterior distribution $\pi(x_i|\boldsymbol{y})$.

If a complex process is under study, it is sometimes easier to model it by thinking of its mechanisms in a hierarchy of underlying processes or distributions (Banerjee et al., 2004). The annual runoff $\boldsymbol{x}$ can e.g. be thought of as a process that depends on some parameters $\boldsymbol{\theta}$ that express the spatial correlation between locations. Here, both $\boldsymbol{x}$ and $\boldsymbol{\theta}$ are stochastic variables with prior (and posterior) distributions. A Bayesian model of this type is typically expressed as a three-staged hierarchical model where the first stage consists of the observation likelihood $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$, the second stage is the prior distribution $\pi(\boldsymbol{x}|\boldsymbol{\theta})$, often referred to as the latent model or process model, while the third stage is the prior distribution of the model parameters $\pi(\boldsymbol{\theta})$. As before, Bayes' formula can be used to make inference about the variables of interest $\boldsymbol{x}$, but also about the model parameters $\boldsymbol{\theta}$ given the set of observations $\boldsymbol{y}$. In this study we use a three-staged hierarchical Bayesian model to model annual runoff.

3.2 Gaussian random fields

Gaussian random fields (GRFs) are commonly used to model environmental variables like precipitation, runoff and temperature or other phenomena

that are continuous in space and/or time. In this analysis, the second stage of the Bayesian hierarchical model consists of GRFs that model the spatial dependency of runoff between catchments. A continuous field $\{x(\boldsymbol{u}); \boldsymbol{u} \in \mathcal{D}\}$ defined on a spatial domain $\mathcal{D} \in \mathcal{R}^2$ is a GRF if for any collection of locations $\boldsymbol{u}_1, ..., \boldsymbol{u}_n \in \mathcal{D}$ the vector $(x(\boldsymbol{u}_1), ..., x(\boldsymbol{u}_n))^{\mathrm{T}}$ follows a multivariate normal distribution (Cressie, 1993), i.e. $(x(\boldsymbol{u}_1), ..., x(\boldsymbol{u}_n))^{\mathrm{T}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}$ is a vector of expected values and $\boldsymbol{\Sigma}$ is the covariance matrix. The covariance matrix $\boldsymbol{\Sigma}$ defines the dependency structure in the spatial domain, and element (i, j) is typically constructed from a covariance function $C(\boldsymbol{u}_i, \boldsymbol{u}_j)$. The dependency structure for a spatial process is often characterized by two parameters: The marginal variance σ^2 and the range ρ. The marginal variance provides information about the spatial variability of the process of interest, while the range gives information about how the covariance between the process at two locations decays with distance. The range is defined as the distance at which the correlation between two locations in space has dropped to almost 0. If the range and the marginal variance are constant over the spatial domain, we have a stationary GRF.

In this study, the involved GRFs have their dependency structure specified by a stationary Matérn covariance function that is given by

$$C(\boldsymbol{u}_i, \boldsymbol{u}_j) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\kappa||\boldsymbol{u}_j - \boldsymbol{u}_i||)^{\nu} K_{\nu}(\kappa||\boldsymbol{u}_j - \boldsymbol{u}_i||). \tag{2}$$

Here, $||\boldsymbol{u}_j - \boldsymbol{u}_i||$ is the Euclidean distance between two locations $\boldsymbol{u}_i, \boldsymbol{u}_j \in \mathcal{R}^d$, K_{ν} is the modified Bessel function of the second kind and order $\nu > 0$, $\Gamma(\cdot)$ is the gamma function and σ^2 is the marginal variance that controls the spatial variability (Guttorp and Gneiting, 2006). The parameter κ is the scale parameter, and it can be shown empirically that the spatial range can be expressed as $\rho = \sqrt{8\nu}/\kappa$, where ρ is defined as the distance at which the correlation between two locations has dropped to 0.1. Using a Matérn GRF is convenient for computational reasons because it makes it possible to use the SPDE approach to spatial modeling from Lindgren et al. (2011) which is briefly described in Section 4.3.

3.3 Kriging and Top-Kriging

Within the geostatistical framework, Kriging approaches have shown promising results for interpolation of hydrological variables (see e.g. Gottschalk (1993), Sauquet et al. (2000) or Merz and Blöschl (2005)). In Kriging methods, the target variable is represented as a random field, typically a Gaussian random field $x(\boldsymbol{u})$ defined through a covariance structure and some unknown parameters. The process of interest is observed at n locations $\boldsymbol{u}_1, ..., \boldsymbol{u}_n$, and any unknown parameters can be estimated based on e.g. maximum likelihood procedures. Furthermore, to estimate the value of the variable $\hat{x}(\boldsymbol{u}_0)$ at an unobserved location $\boldsymbol{u}_0$ a weighted average of the observations is used, i.e.

$$\hat{x}(\boldsymbol{u_0}) = \sum_{i=1}^{n} \lambda_i x(\boldsymbol{u}_i), \tag{3}$$

where λ_i are interpolation weights and $x(\boldsymbol{u}_i)$ for $i = 1, ..n$ are observed values. The interpolation weights are computed by assuming that $\hat{x}(\boldsymbol{u}_0)$ is the Best Linear Unbiased Estimator (BLUE) of $x(\boldsymbol{u}_0)$. That is, we determine $\hat{x}(\boldsymbol{u}_0)$ by finding the weights that both minimize the mean squared error and that give zero mean expected error (Cressie, 1993). Mark that the consequence of the latter, is that the Kriging weights are restricted to sum to 1, i.e. $\sum_{i=1}^{n} \lambda_i = 1$ if we assume that the process is homogeneous in space.

Further, to minimize the mean squared error of the Kriging-predictor in Equation (3), the covariance function (or variogram) must be estimated and evaluated. The covariance function typically depends on the distance between the observations and the target locations, such that observations measured close to the target location $\boldsymbol{u}_0$ are weighted more than observations further away. In many hydrological applications, the centroids of the catchments are used to compute the catchment distances (Merz and Blöschl, 2005; Skøien et al., 2003), but as mentioned in the introduction this can lead to a violation of basic mass conservation laws. The reason is that streamflow variables are connected to (catchment) areas, not single point locations. Catchments are also organized into subcatchments, and this should be considered when computing the Kriging weights.

The Top-Kriging approach suggested by Skøien et al. (2006) is an example of a method that takes the nested structure of catchments into account. In this method, the streamflow observations are interpreted as areal referenced, and the covariance is computed based on the pairwise distances between all grid nodes in a discretization of the involved catchments. This way, observations from a subcatchment can be weighted more than observations from nearby, non-overlapping catchments. Top-Kriging is currently one of the leading methods for interpolation of hydrological variables (Viglione et al., 2013) and is therefore chosen as a benchmark when we evaluate our new interpolation approach.

3.4 Methods for exploiting short records (record augmentation techniques)

The framework we suggest is both a framework for spatial interpolation and a framework for record augmentation. There exist several approaches for record augmentation for which many of them are based on developing a linear relationship between the target catchment and one or several catchments with longer time series of runoff (Fiering, 1963; Laaha and Blöschl, 2005; Matalas and Jacobs, 1964). One class of approaches is the maintenance of variance extension (MOVE) methods. MOVE methods are based on developing a linear relationship between the target catchment and the donor(s) catchment(s) by assuming that the sample mean and sample variance of runoff are maintained over time for the target catchment (Hirsch, 1982). There are different ways the sample mean and sample variance can be estimated, giving different estimators for the predicted runoff. Another way to develop a linear relationship between a donor and a target catchment, is to use simple linear regression (Hirsch, 1982). In this article, we use simple linear regression as a benchmark method, in addition to Top-Kriging.

Assume annual runoff is observed for year $1, ..., n$ in the target catchment and that there exist annual runoff data from some other catchments for year $1, ..., n + m$. Simple linear regression is performed by first finding a so-called donor catchment for the catchment of interest. This can be e.g. the closest catchment in space or a catchment with similar catchment

characteristics (elevation, annual precipitation, vegetation). Next, it is assumed that there is a linear relationship between the annual runoff in the target catchment and the donor catchment, $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ for $i = 1...n$, where y_i is the the annual runoff in the target catchment, x_i is the annual runoff in the donor catchment, ϵ_i is normal distributed measurement error $\mathcal{N}(0, \sigma^2)$ with fixed (but typically unknown) variance σ^2, and β_0 and β_1 are coefficients that must be estimated. The linear relationship between the two catchments is developed by estimating β_0 and β_1 by minimizing the sum of least squares, $\sum_{i=1}^{n}(y_i - (\beta_0 + \beta x_i))^2$. Next, the linear relationship can be used to estimate the runoff at the target catchment $y_{n+1}, ..., y_{n+m}$ based on $x_{n+1}, ..., x_{n+m}$ with corresponding uncertainty estimates.

4 A geostatistical framework for exploiting long-term averages and short records

In this section we present the suggested Bayesian geostatistical framework for runoff interpolation. We start by developing a three staged hierarchical model for annual runoff consisting of a process model, an observation likelihood and prior distributions as described in Section 3.1. Next, we highlight two model properties that make the suggested framework different from most other methods used for interpolation in hydrology (Section 4.2) and explain how the framework is made computationally feasible (Section 4.3).

4.1 Hierarchical model for annual runoff

4.1.1 True annual runoff (process models)

Let the spatial process $\{q_j(\boldsymbol{u}) : \boldsymbol{u} \in \mathcal{D}\}$ denote the runoff generating process at a point location $\boldsymbol{u}$ in the spatial domain $\mathcal{D} \in \mathcal{R}^2$ in year j. The

true annual runoff generated at point location $\boldsymbol{u}$ in year j is modeled as

$$
\begin{aligned}
&q_j(\boldsymbol{u}) = \beta_c + c(\boldsymbol{u}) + \beta_j + x_j(\boldsymbol{u}) \qquad j = 1, .., r, \\
&\pi(\beta_c) \sim \mathcal{N}(0, (10000 \text{ mm/year })^2); \\
&\pi(\beta_j|\sigma_\beta) \sim \mathcal{N}(0, \sigma_\beta^2) \\
&\pi(c(\boldsymbol{u})|\rho_c, \sigma_c) \sim \text{GRF}(\rho_c, \sigma_c) \\
&\pi(x_j(\boldsymbol{u})|\rho_x, \sigma_x) \sim \text{GRF}(\rho_x, \sigma_x)
\end{aligned} \tag{4}
$$

where β_c is an intercept common for all years $j = 1, ..r$ that models the average runoff in the study area over time, while β_j is a year specific intercept that models the annual discrepancy from the long-term average runoff. Likewise is $c(\boldsymbol{u})$ a spatial effect that models the long-term spatial variability of runoff that is caused by climatic conditions in the study area, while $x_j(\boldsymbol{u})$ is a year specific spatial effect that models the spatial variability due to annual discrepancy from the climate. We emphasize that in this context, climate is for simplicity used as a collective term that describes both runoff generation caused by long-term weather-patterns *and* the runoff generation due to catchment characteristics like e.g. elevation and slope. The two spatial effects are modeled as Gaussian random fields (GRFs) with zero mean and stationary Matérn covariance functions with $\nu = 1$, given a range and a marginal variance parameter; $c(\boldsymbol{u})$ with range parameter ρ_c and marginal variance σ_c^2, and $x_j(\boldsymbol{u})$ with range parameter ρ_x and marginal variance σ_x^2. Furthermore, the spatial fields $x_j(\boldsymbol{u})$ for $j = 1, .., r$ are assumed to be independent realizations, or replicates, of the same underlying field to increase the identifiability of the model parameters (Ingebrigtsen et al., 2015). The same applies for the year-dependent intercepts β_j that are all assigned a Gaussian prior $\mathcal{N}(0, \sigma_\beta^2)$ given the variance parameter σ_β^2. The intercept β_c is assigned the weakly informative wide Gaussian prior $\mathcal{N}(0, (10000 \text{ mm/year })^2)$.

So far, runoff has been defined for point locations in space. However, runoff observations are linked to catchment areas, and we need to define the true average annual runoff generated inside a catchment $\mathcal{A}$. We suggest two alternative models: The first model is denoted **the areal model**. For the areal model, the true annual runoff in catchment $\mathcal{A}$ in year j is given by

the average point runoff over the catchment area, i.e.

$$Q_j(\mathcal{A}) = \frac{1}{|\mathcal{A}|} \int_{\boldsymbol{u} \in \mathcal{A}} q_j(\boldsymbol{u}) d\boldsymbol{u}, \tag{5}$$

where $|\mathcal{A}|$ is the catchment area and $q_j(\boldsymbol{u})$ is the point runoff from Equation (4). Interpreting annual runoff as an integral of point runoff ensures that the water balance is approximately preserved for the posterior mean runoff for any point in the landscape. Thus, the areal model is a model for mass-conserved hydrological variables. It also gives a realistic representation of distances and hence also the correlation between the catchments under study (see Equation (2)).

The second model for the annual runoff generated inside a catchment area is denoted **the centroid model**. For the centroid model, the true average annual runoff inside a catchment $\mathcal{A}$ in year j is given by

$$Q_j(\mathcal{A}) = q_j(\boldsymbol{u}_{\mathcal{A}}), \tag{6}$$

where $q_j(\boldsymbol{u}_{\mathcal{A}})$ is the point runoff from Equation (4), and $\boldsymbol{u}_{\mathcal{A}}$ is the centroid of catchment $\mathcal{A}$. This alternative does not provide a preservation of the water balance for the posterior mean predicted runoff and can be used for any point referenced environmental variable. Distances are measured between catchment centroids, such that this method is more similar to the traditional Kriging-methods described in Section 3.3.

4.1.2 Observation likelihood

The true annual runoff from Section 4.1.1 is observed with uncertainty through streamflow data from n catchments which we denote $\mathcal{A}_1, ..., \mathcal{A}_n$. We use the following model for the observed runoff y_{ij} in catchment $\mathcal{A}_i$ in year j

$$\begin{aligned} &y_{ij} = Q_j(\mathcal{A}_i) + \epsilon_{ij}; \quad i = 1, ..n, \quad j = 1, .., r. \\ &\pi(y_{ij}|\sigma_y) \sim \mathcal{N}(Q_j(\mathcal{A}_i), s_{ij}\sigma_y^2). \end{aligned} \tag{7}$$

Here, $Q_j(\mathcal{A}_i)$ is the true runoff from Equation (5) if we use the areal model, or the true runoff from Equation (6) if we use the centroid model. The

error terms ϵ_{ij} are identically, independently distributed as $\mathcal{N}(0, s_{ij}\sigma_y^2)$ given the parameter σ_y^2, and we assume that each observation has its own uncertainty by scaling the variance parameter σ_y^2 with a fixed factor s_{ij} that is further specified in Section 4.1.3.

Through the observation likelihood and the areal formulation of annual runoff from Equation (5), the areal model puts (soft) constraints on the annual runoff over the catchment areas of the gauged catchments. This way the areal model is able to influence the model to distribute the observed annual runoff within the catchment areas and not only at certain gauging points which is what the centroid model does. This represents a potential benefit for the areal model compared to the centroid model when modeling runoff. However, imposing constraints on areas also comes with a computational cost.

4.1.3 Prior models

According to the model specification in Section 4.1.1 and 4.1.2, there are 6 model parameters in the suggested hierarchical model for annual runoff, i.e. $(\sigma_y, \rho_c, \sigma_c, \rho_x, \sigma_x, \sigma_\beta)$. As we apply the Bayesian framework, these have to be given prior distributions, and we use knowledge based priors for most parameters. Note that since the priors are based on expert opinions about the study area, they are specific for the Norwegian dataset and should be modified before further use for other countries or environmental variables.

In the observation model for runoff in Equation (7), each observation is allowed to have its own measurement uncertainty by scaling the variance parameter σ_y^2, with a fixed scale s_{ij}. This makes sense because the spatial variability of mean annual runoff in Norway is large, with values ranging from around 400 mm/year to 4000 mm/year, and heteroscedastic errors can be expected (Petersen-Øverleir, 2004). In the specification of the prior standard deviation $\sqrt{\sigma_y^2 s_{ij}}$, we assume that the measurement uncertainty for runoff increases with the magnitude of the observed value y_{ij}. Based on this we suggest the following scaling factors:

$$s_{ij} = (0.025 \cdot y_{ij}/1000)^2, \tag{8}$$

where y_{ij} is the observed runoff in catchment i in year j in mm/year. The scaling factors are chosen based on what the data provider NVE believes are realistic standard deviations for the observed values, around 2.5% of the observed runoff. They are scaled down by 1000 to achieve appropriate values for $s_{ij}\sigma_y^2$. For the variance parameter σ_y^2, we use the penalized complexity prior (PC prior) suggested by Simpson et al. (2017). The PC prior is a prior constructed for the precision, i.e. the inverse of the variance, and the PC prior for the precision τ of a Gaussian effect $\mathcal{N}(0, \tau^{-1})$ has density

$$\pi(\tau) = \frac{\lambda}{2}\tau^{-3/2}\exp(-\lambda\tau^{-1/2}), \qquad \tau > 0, \quad \lambda > 0, \tag{9}$$

where λ is a parameter that determines the penalty of deviating from a simpler base model. The parameter λ can be specified through a quantile u and a probability α by $\text{Prob}(\sigma > u) = \alpha$, where $u > 0$, $0 < \alpha < 1$ and $\lambda = -\ln(\alpha)/u$. Here, $\sigma = 1/\sqrt{\tau}$ is the standard deviation of this Gaussian distribution. In our case, we specify the PC prior for σ_y as

$$\text{Prob}(\sigma_y > 1500 \text{ mm/year}) = 0.1. \tag{10}$$

Recall that σ_y is scaled with s_{ij} in the final uncertainty model such that a prior 95 % credible interval for the standard deviation $\sqrt{(\sigma_y^2 s_{ij})}$ for the observed runoff in catchment $\mathcal{A}_i$ year j becomes $(0.04, 6)\%$ of the observed value y_{ij}. This is a quite strict prior that is chosen in order to influence the posterior observation uncertainty to be as low as possible. The reason behind this modeling choice is further described in Section 4.2. However, an observation uncertainty of 0.04-6 % of the observed value also corresponds quite well to what NVE knows about the measurement uncertainty for runoff in the study area. Percentages around 2.5% are as mentioned realistic.

For the spatial ranges ρ_x and ρ_c and the marginal variances σ_x^2 and σ_c^2 for the Gaussian random fields $x_j(\boldsymbol{u})$ and $c(\boldsymbol{u})$, we use the joint informative PC prior suggested in Fuglstad et al. (2019). It is specified through the following probabilities and quantiles:

$$\begin{aligned} &\text{Prob}(\rho_x < 20 \text{ km}) = 0.1, \quad &&\text{Prob}(\sigma_x > 2000 \text{ mm/year}) = 0.1, \\ &\text{Prob}(\rho_c < 20 \text{ km}) = 0.1, \quad &&\text{Prob}(\sigma_c > 2000 \text{ mm/year }) = 0.1. \end{aligned}$$

The percentages and quantiles are chosen based on expert knowledge about the spatial variability in the area of interest. It is reasonable to assume that locations that are less than 20 km apart are correlated when it comes to runoff generation. In Norway the annual runoff varies from around 300 mm/year - 6000 mm/year such that a marginal standard deviation that is below 2000 mm/year is reasonable. The parameters of the climatic GRF $c(\boldsymbol{u})$ and the year dependent GRF $x_j(\boldsymbol{u})$ are given the same prior as it is difficult to identify if the spatial variability mainly comes from climatic processes or from annual variations. We also want the data to decide which of the two effects that dominates in the study area, and in this way detect hydrological spatial stability or instability. Recall that the phrase *hydrological spatial stability* here is used to describe a variable and/or a study area that is characterized by an underlying spatial pattern that is repeated over time.

As specified in Section 4.1.1, the year specific intercepts β_j for $j = 1, .., r$ are all assigned the same Gaussian prior $\mathcal{N}(0, \sigma_\beta^2)$ given the standard deviation parameter σ_β. The standard deviation σ_β is given the PC prior from Equation (9) specified by the wide prior $P(\sigma_\beta > 10 \text{ m/year}) = 0.2$. With this prior, the prior 95% credible interval is approximately $(0.002, 40.5)$ m/year for the standard deviation σ_β of β_j.

4.1.4 Feasible computation of catchment runoff for the areal model

In the areal model in Equation (5), the true runoff is modeled as the integral of point runoff over a catchment. To make the areal model computationally feasible, the integral is calculated by a finite sum over a discretization of the target catchment. More specifically, if $\mathcal{L}_i$ denote the discretization of catchment $\mathcal{A}_i$, the annual runoff in catchment $\mathcal{A}_i$ in year j is calculated as

$$Q_j(\mathcal{A}_i) = \frac{1}{N_i} \sum_{\boldsymbol{u} \in \mathcal{L}_i} q_j(\boldsymbol{u}), \tag{11}$$

where N_i is the number of grid nodes in the discretization $\mathcal{L}_i$. In the discretization of the catchments it is important that a subcatchment shares

grid nodes with its overlapping catchment(s) such that the water balance can be preserved. In our analysis, we use a regular grid with 4 km spacing. It is also important that the discretization of the study area is fine enough to capture the rapid changes of annual runoff in the study area. Otherwise, non-realistic results such as negative runoff can occur.

4.1.5 Full model specification

Assuming that we observe runoff at n stream gauges for $j = 1, ..r$ years and that $\mathcal{L}_{\mathcal{D}}$ contains all grid nodes in the discretization of the catchments $\mathcal{L}_{\mathcal{A}_i}$ for $i = 1, ..., n$, the areal model in Section 4.1.1 - 4.1.4 can be summarized as the following hierarchical geostatistical model:

$$\pi(\boldsymbol{y}|\boldsymbol{x}, \sigma_y) \sim \prod_{j=1}^{r}\prod_{i=1}^{n}(I\{\text{Observation } \ y_{ij} \text{ is available}\} \cdot \mathcal{N}(Q_j(\mathcal{A}_i), s_{ij}\sigma_y^2) + 1 \cdot I\{\text{Observation } \ y_{ij} \ \text{ is } \ \text{missing}\}) \quad \text{[Observation likelihood]}$$

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}) = \pi\left(c(\boldsymbol{u_1}), ..., c(\boldsymbol{u_m})|\rho_c, \sigma_c\right) \cdot \pi(\beta_c) \cdot \prod_{j=1}^{r}[\pi\left(x_j(\boldsymbol{u_1}), ..., x_j(\boldsymbol{u_m})|\rho_x, \sigma_x\right) \cdot \pi(\beta_j|\sigma_\beta)] \quad \text{[Latent Model]} \tag{12}$$

$$\pi(\sigma_y, \boldsymbol{\theta}) = \pi(\rho_x, \sigma_x) \cdot \pi(\rho_c, \sigma_c) \cdot \pi(\sigma_\beta) \cdot \pi(\sigma_y) \quad \text{[Prior]}$$

where $\boldsymbol{y}$ is a vector containing all runoff observations y_{ij} from all catchments i and years j, $\boldsymbol{x}$ is a vector containing all latent variables, i.e. the intercepts β_c, β_j and the GRFs $c(\boldsymbol{u}.)$ and $x_j(\boldsymbol{u}.)$ for all combinations of grid nodes $\boldsymbol{u_1}, ..., \boldsymbol{u_m} \in \mathcal{L}_{\mathcal{D}}$ and years j=1,..,r. Furthermore, $Q_j(\mathcal{A}_i)$ is the true annual runoff that is modeled as a function of the latent field $\boldsymbol{x}$, while $I(\cdot)$ is an indicator function that is equal to 1 if its argument is true, and 0 otherwise allowing for missing data and short records of runoff. Finally, $\boldsymbol{\theta} = (\rho_x, \sigma_x, \rho_c, \sigma_c, \sigma_\beta)$. Together with σ_y it contains all model parameters.

The centroid model is summarized as a hierarchical model similarly, except that the true annual runoff $Q_j(\mathcal{A}_i)$ is given by Equation (6) instead of

Equation (11). This also means that the grid nodes $\boldsymbol{u_1}, ..., \boldsymbol{u_m}$ in the above hierarchical model must be replaced by $\boldsymbol{u}_{\mathcal{A}_1}, ..., \boldsymbol{u}_{\mathcal{A}_n}$, i.e. the locations of the centroids of the n catchments under study.

The purpose of Bayesian inference is to estimate the posterior distributions of the latent variables $\boldsymbol{x}$ and the parameters $\boldsymbol{\theta}$ based on the observations $\boldsymbol{y}$ as described in Section 3.1. In this study, the resulting distributions are used to quantify the variable of interest, the catchment runoff $Q_j(\mathcal{A})$. By Equation (6) and Equation (11) we see that the catchment runoff is determined by the point runoff $q_j(\boldsymbol{u_1}), .., q_j(\boldsymbol{u_m})$ which is again determined by the latent field $\boldsymbol{x}$ through Equation (4). This means that in the process of estimating the catchment runoff $Q_j(\mathcal{A})$ we always estimate the point runoff $q_j(\boldsymbol{u})$ and the latent field $\boldsymbol{x}$ first. To clarify this process, consider Figure 10 that is presented later in the article. This shows the posterior mean runoff $q_j(\boldsymbol{u})$, or $\pi(\boldsymbol{x}|\boldsymbol{y})$ implicitly, for all points in the study area. From these point estimates, predictions for the *areal* model $Q_j(\mathcal{A})$ are obtained by taking the average of $q_j(\boldsymbol{u})$ over relevant grid nodes according to Equation (11). For the centroid model, a catchment areal prediction $Q_j(\mathcal{A})$ is obtained by simply extracting the value of $q_j(\boldsymbol{u}_{\mathcal{A}})$ at the catchment centroid $\boldsymbol{u}_{\mathcal{A}}$ according to Equation (6). From the point referenced predictions in Figure 10 we this way obtain catchment predictions like the ones presented later in e.g. Figure 7.

From the hierarchical formulation in (12) we also note that the framework takes the time dimension into account through multiplying the likelihood for annual runoff over different years $j = 1, ..r$. These years don't need to be consecutive, which allows for e.g. combining old measurements from closed stations with more recent data. Different years of data are connected through the constant climatic component $(c(\boldsymbol{u}) + \beta_c)$. Apart from this, there is no temporal dependency in the model that assumes correlation over time, and routing is not taken into account. This makes sense for our suggested application, as there is no prominent time dependency for annual runoff in Norway (see e.g. Figure 2b). Routing effects can typically be neglected for time-aggregated runoff variables for longer time scales. For shorter time scales for which routing has an impact, other spatio-temporal models should be considered, for example the one in Skøien and Blöschl (2007).

4.2 Two model properties and contributions

In this section we highlight and describe two of the model properties that make the suggested framework different from Top-Kriging and geostatistical interpolation methods that are typically used for hydrological applications.

4.2.1 Exploiting short records

The first property we highlight is how the model is particularly suitable for exploiting short records of runoff, and this holds for both the areal model and the centroid model. This property is already briefly addressed in the introduction, and is enabled because we simultaneously model several years of data with a spatial component $c(\boldsymbol{u})$ that is common for all years under study. The GRF $c(\boldsymbol{u})$ represents the long-term spatial variability of runoff. If most of the spatial variability can be explained by long-term patterns, the marginal variance parameter σ_c^2 will dominate over the marginal variance parameter σ_x^2 of the annual GRF $x_j(\boldsymbol{u})$ (and the other model variances), i.e. $\sigma_c \gg \sigma_x$. Thus, a short record of runoff from an otherwise ungauged catchment will have a large impact also for predictions in years without data through $c(\boldsymbol{u})$. On the other hand, if most of the annual runoff is explained by year specific effects, $x_j(\boldsymbol{u})$ will dominate over $c(\boldsymbol{u})$ and short records will not have a large impact on the final model. Hence, it is safe to include short records in the model regardless of the weather-patterns in the study area.

Existing methods for exploiting short records are typically based on linear regression or computing the correlation between the runoff in the target catchment and one or several donor catchments, and in order to perform these procedures the short record must be of length larger than one (Fiering, 1963; Laaha and Blöschl, 2005). In the method we suggest, it is possible to include a short record of length one, and it is already shown for a smaller case study that this often is enough to see a large improvement in the predictability of (annual) runoff for certain climates (Roksvåg et al., 2020).

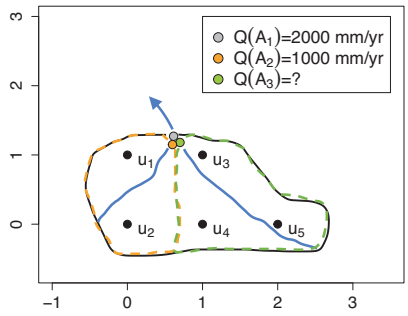


Figure 6: Conceptual, simplified figure of a river network, for which the involved catchments are discretized by 5 grid nodes $\boldsymbol{u}_1, .., \boldsymbol{u}_5$, and each grid node represent one areal unit. Catchment $\mathcal{A}_1$ contains all grid nodes $\boldsymbol{u}_1, .., \boldsymbol{u}_5$, Catchment $\mathcal{A}_2$ consists of grid nodes $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, while Catchment $\mathcal{A}_3$ consists of grid nodes $\boldsymbol{u}_3, \boldsymbol{u}_4$ and $\boldsymbol{u}_5$. Hence, this is a system of nested catchments where $\mathcal{A}_1$ covers $\mathcal{A}_2$ and $\mathcal{A}_3$. Assume that there are observations of annual runoff at the outlet of catchment $\mathcal{A}_1$ and catchment $\mathcal{A}_2$: $Q(\mathcal{A}_1)$=2000 mm/year and Q($\mathcal{A}_2$)=1000 mm/year. Catchment $\mathcal{A}_3$ is ungauged. In order to fulfill water balance constraints of the areal model from Equation (11), imposed by the likelihood in Equation (7), the predicted mean annual runoff in catchment $\mathcal{A}_3$ must be around 2667 mm/year if we assume a low observation uncertainty.

4.2.2 The water balance constraints of the areal model

The second property we highlight only holds for the areal model, and is related to its mass-conserving properties and its ability to do more than smoothing: Runoff is in Equation (11) defined as an weighted sum of point runoff. Through Equation (11) and the likelihood defined in Equation (7), a (soft) constraint is put on the predicted annual runoff for the catchments for which we have observations. This also has the consequence that the suggested model allows us to predict values that are larger than any of the observed values in the area of interest. As a conceptual example, consider the river network in Figure 6, where each black node represents one areal unit in the discretization of the catchments. The observed runoff is 1000 mm/year in the subcatchment and 2000 mm/year in the surrounding larger catchment. That means that the constraints imposed by the observation likelihood and Equation (11) are the following:

$$
\begin{aligned}
1000 \text{ mm/year} &= (q(\boldsymbol{u}_1) + q(\boldsymbol{u}_2))/2 + \text{uncertainty} \\
2000 \text{ mm /year} &= (q(\boldsymbol{u}_1) + q(\boldsymbol{u}_2) + q(\boldsymbol{u}_3) + q(\boldsymbol{u}_4) + q(\boldsymbol{u}_5))/5 + \text{uncertainty}.
\end{aligned} \tag{13}
$$

As described in Section 4.1.3 we use a quite strict prior on the observation uncertainty. This is done to try to force the above soft constraints to be stronger. By solving this system of equations, it can be shown that the predicted value in the ungauged catchment $\mathcal{A}_3$ must be

$$
\hat{Q}(\mathcal{A}_3) = \frac{q(\boldsymbol{u}_3) + q(\boldsymbol{u}_4) + q(\boldsymbol{u}_5)}{3} + \text{uncertainty} = 2667 \text{ mm/year} + \text{uncertainty}
$$

where the above uncertainty term is determined by the observation uncertainties for a fixed set of ranges and marginal variances. Hence, as long as the observation uncertainty is sufficiently low, the predicted runoff in the unobserved area $\mathcal{A}_3$ in Figure 6 is close to 2667 mm/year which is larger than any of the observed values. This example illustrates that the areal model is able to go beyond smoothing through its runoff constraints. This makes the model different from many other interpolation methods, that only rely on spatial smoothing. For the above example, such methods would typically produce a prediction between 1000 mm/year and 2000 mm/year.

The full areal model is of course slightly more complicated than the simple example above, as prior distributions, covariance calculations and spatial ranges must be taken into account. However the simple example illustrates the general idea of how the observation likelihood interprets the areal observations and constrains runoff. That the full areal model actually is able to conserve mass in practice, is demonstrated for a real case example from Norway in Roksvåg et al. (2020).

The constraints in Equation (13) also show how the areal model ensures consistent predictions over nested catchments: As the predicted runoff in the main catchment $\mathcal{A}_1$ can be expressed as a weighted sum of the predicted runoff in all its subcatchments depending on catchment areas, i.e. as $\hat{Q}(\mathcal{A}_1) = \frac{2}{5}\hat{Q}(\mathcal{A}_2) + \frac{3}{5}\hat{Q}(\mathcal{A}_3)$, the water balance can not be violated for the predicted runoff for any of the catchments in Figure 6. This means that the equations in (13) correspond to water balance constraints.

Compared to Top-Kriging, both Top-Kriging and the proposed method assume that the underlying variable is linearly aggregated and mass-conserved when performing covariance calculations. Top-Kriging is also able to predict larger values than any of the observed values by allowing negative Kriging weights. However, Top-Kriging does not use constraints to ensure that the mass balance is preserved over nested catchments, as in the above example. Consequently, the Top-Kriging predictions can more easily violate the water balance, which can have both benefits and drawbacks depending on the target variable and the problem we are trying to solve. Another hydrological model that uses water balance constraints, not unlike the proposed method, is the Kriging approach in Sauquet et al. (2000) where mass balance constraints are introduced as additional constraints in the Kriging system of equations.

4.3 Inference

In order to make the framework described in Section 4 computationally feasible, some simplifications of the suggested models are necessary. In general, statistical inference on models including GRFs is slow when the number of target locations is large because matrix operations on dense covariance matrices are required. The computational complexity is particularly large for the areal model, because each grid node in the discretization of the catchments can be regarded as a new target location, and because it includes soft constraints. To solve the computational issues for the centroid and areal model, we utilize that a GRF with a Matérn covariance function can be expressed as the solution of a specific Stochastic partial differential equation (SPDE) (Lindgren et al., 2011). This SPDE can be solved by using the finite element method (see e.g. Brenner and Scott (2008)), and the result is a Gaussian Markov random field (GMRF). Working with GMRFs is convenient because GMRFs have precision matrices (inverse covariance matrices) that typically are sparse with more zero elements, and efficient algorithms are available for sparse matrix operations (see e.g Rue and Held (2005)). In this work, both GRFs $x_j(\boldsymbol{u})$ and $c(\boldsymbol{u})$ are approximated by GMRFs.

Another challenge with the suggested models, is that we suggest Bayesian

models that include a large number of parameters for which the marginal distributions must be estimated. Traditionally, Bayesian inference is done by using Markov chain Monte Carlo-methods (MCMC), but inference can be slow when the dimension of the problem is large (Gamerman and Lopes, 2006). These challenges are met by modeling runoff as a latent Gaussian model (LGM). That is, the latent part $\boldsymbol{x}$ of the hierarchical model in 4.1.5 consists of only Gaussian distributions. More specifically, the prior distributions for $c(\boldsymbol{u})$ and $x_j(\boldsymbol{u})$ are modeled as GRFs, and the prior distributions for β_j and β_c are Gaussian given the model parameters (see the equations in (4)). This is convenient, because it allows us to use integrated nested Laplace approximations (INLA) to make inference and predictions. INLA is a tool for making Bayesian inference for LGMs (Rue et al., 2009) and represents a fast and approximate alternative to MCMC algoithms. The INLA approach is based on approximating the marginal distributions by using Laplace or other analytic approximations, and on numerical integration schemes. The main computational tool is the sparse matrix calculations described in Rue and Held (2005), such that in order to work fast, the latent field of the LGM should be a GMRF with a sparse precision matrix. This requirement is fulfilled through the SPDE approach as already outlined.

INLA in general provides approximations of very high accuracy for most models (Rue et al., 2009; Martino et al., 2011; Eidsvik et al., 2012; Huang et al., 2017), but has faced problems for some (more extreme) models with binomial or Poisson data (Fong et al., 2009; Ferkingstad and Rue, 2015). For Gaussian likelihoods however, INLA is exact up to numerical integration error. As we use Gaussian likelihoods in this work, we can thus expect INLA to give reliable approximations. The SPDE approach also provides accurate approximations (Lindgren et al., 2011; Huang et al., 2017), but it is important that the mesh involved in the finite element method computations is sufficiently dense relative to the spatial variability and range in the study area.

Because of the high computational speed and accuracy, the INLA and SPDE framework has become quite common to use within different fields of science. See for example Khan and Warner (2018); Opitz et al. (2018); Yuan et al. (2017); Guillot et al. (2014); Ingebrigtsen et al. (2014); Bakka

et al. (2018). We refer to the R-package `r-inla` for a user-friendly interface for applying INLA and the SPDE approach to spatial modeling. In particular, Moraga et al. (2017) is recommended for a description of how a model with (catchment) areal data can be implemented in `r-inla`. Furthermore, we have made code for the centroid model available on `http://www.github.com/tjroksva/runoffinterpolation` (doi: 10.5281/zenodo.3630348) with example data from the catchments in Figure 1a.

5 Model evaluation

The main objectives of this article are to (1) evaluate the new framework's ability to fill in missing annual runoff observations and to (2) predict mean annual runoff for catchments with varying record lengths. By this we also want to (3) demonstrate the potential added value of including short runoff records in the modeling compared to not using them. In this section we present the experimental set-up and the evaluation criteria used to address our research questions.

5.1 Experimental set-up for infill of missing annual observations (1996-2005)

To assess the framework's ability to fill in missing values of annual runoff, we do interpolation of runoff for the 10 hydrological years 1996-2005 for the 180 fully gauged catchments shown in Figure 1a. This is done both for series of annual runoff, and for the annual series of monthly runoff for January, April and June described in Section 2.

The annual time series of monthly runoff are included in the analysis in order to demonstrate the framework's properties for hydrological variables or areas that are driven by more unstable hydrological processes. For the annual series of monthly runoff, the models from Section 4 are specified as before: Considering predictions for January, $Q_j(\mathcal{A}_i)$ in Equation (5) represents the true runoff in January for catchment $\mathcal{A}_i$, year j, such that the GRF $c(\boldsymbol{u})$ represents the long-term spatial variability in January. Likewise, the GRF $x_j(\boldsymbol{u})$ represents the annual discrepancy from the climate

in January, and y_{ij} is the observed runoff in January for catchment $\mathcal{A}_i$ year j. The models for June and April are specified similarly, and for simplicity we use the same prior distributions for all experiments.

In our assessment of the framework's predictive performance for infill of missing annual observations, the three following methods are compared:

Top-Kriging: Spatial interpolation with Top-Kriging. For Top-Kriging each year (1996-2005) is interpolated independently from other years. Short records on an annual (or monthly) scale don't have an impact on years without data. The default covariance function (or variogram) in the R package `rtop` was fitted as this gave the most accurate results. This is a multiplication of a modified exponential and fractal variogram model, the same model as used in Skøien et al. (2006).

Areal model: Spatial interpolation with the model defined in Section 4 with true annual runoff given by the areal model from Equation (11). That is, the annual runoff in a catchment is interpreted as the average point runoff over the catchment area. All years are modeled simultaneously (1996-2005) such that short records of data can influence years without data.

Centroid model: Spatial interpolation with the model defined in Section 4 with true annual runoff given by the centroid model from Equation (6). That is, annual runoff is interpreted as a process linked to point locations in space (the catchment centroids), and not to catchment areas. All years are modeled simultaneously (1996-2005) such that short records of data can influence years without data.

The predictive performance of the three methods is evaluated by cross-validation: The 180 catchments in Norway were divided into 20 groups or folds, each containing 9 catchments. In turn each group was left out, and annual or monthly runoff predictions were performed for these so-called target catchments by using observations from the catchments in the other groups. That is, we predict runoff for 1996-2005 for 9 target catchments at once by using data from the remaining 171 fully gauged catchments, and repeat the process for all 20 cross-validation folds. To evaluate and compare the three methods described above, we do the following two tests:

UG (ungauged): Assess the methods' ability to fill in missing values for ungauged catchments (denoted UG). That is, the target catchments are treated as totally ungauged, and all their observations are left out of the dataset when the predictions for 1996-2005 are performed.
PG (partially gauged): Assess the methods' ability to fill in missing values for partially gauged catchments (denoted PG). Each of the 9 target catchments in the cross-validation group is allowed to have one annual observation of runoff. That is, a short record of length one from the target catchment is included in the observation likelihood in addition to the full data series of runoff from the catchments in the other cross-validation folds. The short record is drawn randomly from the ten years of observations available for each target catchment. We perform predictions for 9 partially gauged target catchments at once, for all 10 study years (for which one of them is observed for each catchment), and repeat the process for all 20 cross-validation folds.

To make the results comparable, we use the same cross-validation groups for both experiments (UG and PG) and methods (Top-Kriging, areal model and centroid model), and remove the same set of annual observations for PG across methods. For the PG-case, we also compare our models to a method for exploiting short records from the target catchment. The method we choose for comparison is simple linear regression, and we perform linear regression for the PG-case as follows:
Linear regression: The closest catchment in terms of catchment centroid is used as a donor catchment and only catchments outside the target catchment's cross-validation group can be considered. Two annual observations between 1996 and 2005 are randomly drawn from the target catchment, and data from the donor catchment and target catchment are used to fit a linear regression model on the form $y_i = \beta_1 x_i + \epsilon_i$. Next, the model is fitted as described in Section 3.4, and used to predict runoff for the target catchment for 1996-2005 (where two of the years are observed). The reason for using a short record of length two instead of one, is that at least two observations are required to fit a linear regression model with uncertainty. Also mark that we have omitted the intercept β_0 in the regression model, such that we only have two unknown variables (β_1 and σ^2). We emphasize that estimating the variables statistically based on only two pair of observations

might not give the most reliable estimates, particularly when considering σ^2. However, the linear regression results can provide an intuition about the target variable and how correlated the observation locations are over time. A good performance for linear regression in this study, suggests that the spatial pattern for the target variable is very stable and that a fair prediction can be obtained by simply using the ratio of runoff between a target catchment and a donor catchment for any chosen year to develop a linear relationship between the two catchments. Hence, a short record can be very valuable. Motivated by this, linear regression is treated as an indicator for when our geostatistical method can be expected to perform particularly well, rather than as a recommended method for record augmentation when having only two observations.

5.2 Experimental set-up for predictions of mean annual runoff (1981-2010)

To assess the framework's ability to estimate mean annual runoff, we use annual data from 1981-2010 from the 260 catchments in Figure 5. Recall that these catchments have at least one observation of mean annual runoff between 1981 and 2010, but only 83 of them are fully gauged. This was different from the experiments described in Section 5.1, where all the test catchments were fully gauged before the cross-validation was performed.

For this experiment, we again compare the performances of Top-Kriging, the areal and the centroid model. The areal model and the centroid model are fitted for several years of annual runoff simultaneously, as before. As a predictor for the mean annual runoff, we use the posterior distribution of the climatic part of the model. This is given by $c(\boldsymbol{u}_{\mathcal{A}}) + \beta_c$ for the centroid model, where $\boldsymbol{u}_{\mathcal{A}}$ is the centroid of the catchment $\mathcal{A}$ of interest. For the areal model it is given by the average $c(\boldsymbol{u}_i) + \beta_c$ over the grid nodes $\boldsymbol{u}_i$ in the discretization of the target catchment. Note that the climatic part of the model must be re-estimated for each experiment or cross-validation fold.

In order to interpolate mean annual runoff by using Top-Kriging, we have to compute the mean annual runoff based on the annual observations for all catchments before running the analysis. For catchments with less than 30

annual observations we use the average of the 1-29 available observations as an approximation for the mean annual runoff for 1981-2010. Next, the mean annual runoff is interpolated by using Top-Kriging where the uncertainty of the observations is specified as a function of record length. This is the suggested approach from Skøien et al. (2006) for including short records in the Top-Kriging framework. We set the observation variance for a catchment with record length m to $\hat{\sigma}^2/m$, where $\hat{\sigma}$ is the average empirical standard deviation for the observed annual runoff taken over the 83 fully gauged catchments in our dataset, in this case $\hat{\sigma} = 336$ mm/year. For the Top-Kriging experiments, we fit the same covariance model as in Section 5.1.

The areal and centroid model and Top-Kriging are again evaluated by cross-validation. The 83 fully gauged catchments from Figure 5 were divided into 4 folds containing 20, 20, 20 and 23 catchments respectively, and in turn observations from each fold were removed and predicted. This was done for varying record lengths for the target catchments, more specifically when 0, 1, 3, 5 or 10 randomly drawn annual observations from the target catchments were included in the likelihood. We denote these settings UG, PG1, PG3, PG5 and PG10. Note that while we only are able to assess the predictive performance for the 83 fully gauged catchments in Figure 5, data from the remaining 177 partially gauged catchments in Figure 5 are used in the observation sample. This is in addition to the data from the fully gauged catchments from the other folds.

5.3 Evaluation scores

To evaluate the predictions we use the root mean squared error (RMSE) and the continuous rank probability score (CRPS). Having m pairs of observations and predictions, the RMSE is computed as

$$\text{RMSE} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}(y_j^* - \hat{y}_j^*)^2},$$

where y_j^* is the observed value and $\hat{y}_j^*$ is the corresponding predicted value. In our analysis, the posterior mean is used as a the predicted value for the areal and centroid model.

The CRPS is defined as

$$\mathrm{CRPS}(F, y^*) = \int_{-\infty}^{\infty} (F(s) - 1\{y^* \leq s\})^2 ds,$$

where $F()$ is the predictive cumulative distribution and y^* is the actual observation (Gneiting and Raftery, 2007). For the methods we test (areal, centroid, Top-Kriging and linear regression), $F()$ is a Gaussian distribution with mean equal to the predicted value and standard deviation equal to the standard deviation of the prediction.

For the experiments related to infill of individual years, the CRPS and RMSE are first computed for each of the 180 catchments in the dataset based on 10 pairs of predictions and observations. The average RMSE and CRPS over all catchments are used as a summary scores. For the experiments related to predictions of mean annual runoff, there is only one (mean annual) prediction for each catchment, and the RMSE and CRPS over all catchments are reported. Both the CRPS and the RMSE are negatively oriented such that low scores mean better predictions.

To be able to compare the RMSE and CRPS across methods we use a paired Wilcoxon Signed-Rank Test (Siegel, 1956). This is a non-parametric test that does not require normal distributed data. The null hypothesis of the test is that the median difference between pairs of data (in this case pairs of RMSE or CRPS values) follows a symmetric distribution around zero. The alternative hypothesis is that the difference between the data pairs does not follow a symmetric distribution around zero. If the null hypothesis is rejected, it indicates that one of the methods gives a significantly smaller RMSE or CRPS than another method.

In addition to the RMSE and the CRPS, we report the 95 % coverage of the experiments. The 95 % coverage is computed by calculating the amount of the actually observed runoff values that are within the corresponding 95 % posterior prediction intervals. Here, we make posterior prediction intervals for Top-Kriging and linear regression by assuming that the predictions are Gaussian. A 95 % coverage close to 0.95 is optimal and indicates that the model provides an accurate representation of the underlying uncertainty.

We also want to compare our mean annual runoff results with other studies of mean annual runoff, more specifically the studies collected in Blöschl

et al. (2013). In Blöschl et al. (2013), the absolute normalized error (ANE) and the squared correlation coefficient (r^2) are used as evaluation scores. The ANE is computed as

$$\text{ANE} = \frac{|\hat{y}^* - y^*|}{y^*}, \tag{14}$$

where y^* and $\hat{y}^*$ are the observed and predicted value as before. The ANE divides the absolute difference between the actual observation y^* and corresponding prediction $\hat{y}^*$ by the observed runoff, and is therefore scale independent. An ANE close to zero corresponds to an accurate prediction.

Finally, the squared correlation coefficient between m pairs of observations and predictions is computed as

$$r^2 = (\text{Cor}\{(y_1, ..., y_m), (\hat{y}_1^*, ..., \hat{y}_m^*)\})^2, \tag{15}$$

where $\text{Cor}(\cdot, \cdot)$ denotes the Pearson correlation. An r^2 close to 1 indicates a high correlation between the predicted and observed values.

6 Results

6.1 Predictions for individual years (1996-2005)

We now present the results related to the framework's ability to predict runoff for individual, missing years for the annual time series of annual and monthly runoff for a 10 year period (1996-2005). First, we present the results for the ungauged catchments (UG), before we proceed to the partially gauged catchments (PG) that have short records of length one.

6.1.1 Infill for ungauged catchments (UG)

For the ungauged case (UG), the target catchments are treated as totally ungauged for the ten study years 1996-2005, and missing values are predicted both for annual and monthly runoff. In Figure 7 the resulting

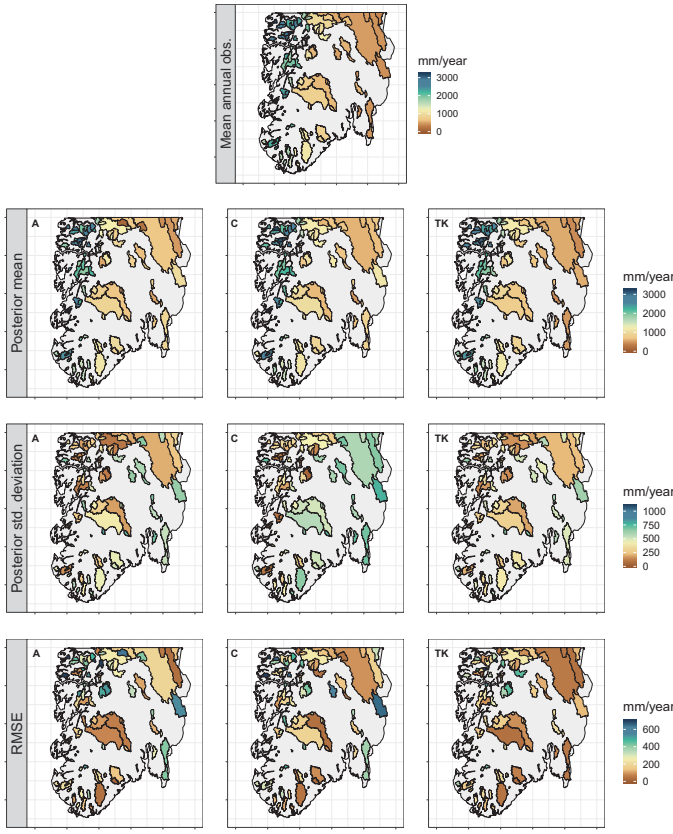


Figure 7: Average posterior mean for $Q_j(\mathcal{A})$, average posterior standard deviation for $Q_j(\mathcal{A})$ and average RMSE for each catchment for predictions of missing annual observations in southern Norway for $j = 1, ..10$ for the areal model (A,left), the centroid model (C, middle) and Top-Kriging (TK, right) when the target catchments are treated as ungauged (UG). The observed mean annual runoff is also included as a reference (first plot).

average predicted *annual* runoff in southern Norway is presented for Top-Kriging, the areal model and the centroid model. The three methods give similar results for the posterior mean, and all are able to reproduce the true spatial pattern of annual runoff. Furthermore, the RMSE plots in Figure 7 show that the three methods succeed and fail for many of the same catch-

ments. Here, we should keep in mind that the RMSE is scale dependent and might not give the best impression of the relative performance across the study area. However, we note that many of the catchments with high RMSE values typically are small catchments located in western Norway. We will come back to these catchments in Section 6.1.2 to see how the predictions here were affected when including a short record.

Table 1: Predictive performance for predictions of missing annual values in ungauged catchments (UG) and partially gauged catchments (PG) for the areal model, centroid model, Top-Kriging (TK) and simple linear regression (LR). The best performance in each row is marked in bold. The RMSE and CRPS were compared across methods by using a one sided paired Wilcoxon Signed-Rank test for assessing the significance of the results. Results that were significantly better than other results are marked with stars.

		RMSE [mm/year]				CRPS [mm/year]				Coverage 95 %			
Case	Dataset	Areal	Centr.	TK	LR	Areal	Centr.	TK	LR	Areal	Centr.	TK	LR
UG	Annual	337	343	**310** *	-	242	249	**225***	-	0.92	0.91	**0.94**	-
UG	January	39	37	**36** *	-	26	25	**24***	-	0.92	0.89	**0.93**	-
UG	April	38	38	**37**	-	25	25	**24**	-	0.89	0.85	**0.93**	-
UG	June	87	96	**82** *	-	59	67	**56** *	-	**0.91**	0.84	**0.91**	-
PG	Annual	**171** **	184 **	290	178 **	**105****	113 **	201	240	**0.95**	0.94	**0.95**	0.96
PG	January	**30****	**30****	33	61	**19****	20**	21	88	0.91	0.89	0.91	**0.95**
PG	April	**31****	33 **	35	50	**20** **	21 **	22	94	0.86	0.84	**0.94**	**0.96**
PG	June	**55** **	63 **	78	95	**35** **	42 **	50	136	0.90	0.84	0.93	**0.96**

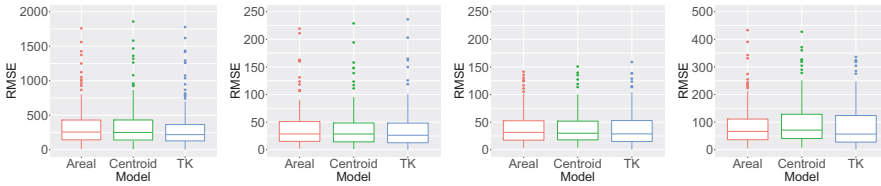
* The RMSE/CRPS is significantly lower than the RMSE/CRPS of the *areal and the centroid model* on a 5 % significance level.

** The RMSE/CRPS is significantly lower than the RMSE/CRPS of *Top-Kriging* on a 5 % significance level.

Considering the posterior standard deviation in Figure 7, we notice that Top-Kriging and the areal model provide a similar quantification of the predictive uncertainty. Top-Kriging and the areal model take the nestedness of catchments into account by interpreting the runoff data as areal referenced, providing a predictive standard deviation of runoff that varies with the size of the target catchment: Figure 7 shows that smaller catchments typically have a larger predictive uncertainty, which is reasonable. For the centroid model, runoff observations are point referenced and weighted independently of catchment size. Consequently, the predictive uncertainty only

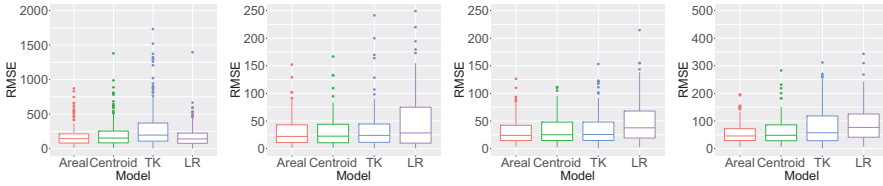
depends on how the centroids of the observed catchments are distributed in space, and decreases in areas where there are clusters of data. The predictive uncertainties provided by Top-Kriging and the areal method are thus more intuitive and realistic considering the process we are studying. The latter is also reflected in the coverage percentages presented in Table 1. The coverages show the amount of the actual observations that were captured by the corresponding 95 % prediction intervals, and these are slightly closer to 0.95 for Top-Kriging and the areal model compared to the centroid model.

Table 1 also presents the summary scores for the predictive performance for infill of missing values for ungauged catchments for all methods. According to the RMSE and CRPS, Top-Kriging is a better interpolation method than our two suggested methods for ungauged catchments. However, the boxplots in Figure 8 illustrate the distribution of RMSE for all catchments, and we see that on a monthly scale, the difference between Top-Kriging and the two other methods is quite low from a practical point of view. For January and April the differences are almost negligible.



(a) Annual (b) January (c) April (d) June

Figure 8: Distribution of RMSE [mm/year and mm/month] for infill of missing values for all catchments and years (1996-2005) when the target catchments are treated as ungauged (UG) in the cross-validation for the areal, centroid and Top-Kriging (TK) method. The lower and upper quartiles correspond to the first and third quartiles (the 25th and 75th percentiles), and the whiskers extend from the quartiles no further than 1.5· IQR, where IQR is the distance between the first and third quartile. The same applies for all boxplots presented in this paper.



(a) Annual (b) January (c) April (d) June

Figure 9: Distribution of RMSE [mm/year and mm/month] for infill of missing values for all catchments and years (1996-2005) for the areal model, centroid model and for Top-Kriging (TK) when the target catchments are treated as partially gauged (PG), i.e. a short record of length one from the target catchment is included in the observation likelihood in the cross-validation. Results for linear regression (LR) are also included here.

6.1.2 Infill for partially gauged catchments (PG)

For the partially gauged (PG) case, each target catchment is allowed to have a short record of length one for Top-Kriging, the areal and centroid model, and length two for linear regression. Before we comment the results from the cross-validation in Table 1 and Figure 9, we consider the posterior estimates of the range parameters (ρ_x and ρ_c) and the marginal variance parameters (σ_x and σ_c) of the year-specific GRF $x_j(\boldsymbol{u})$ and the climatic

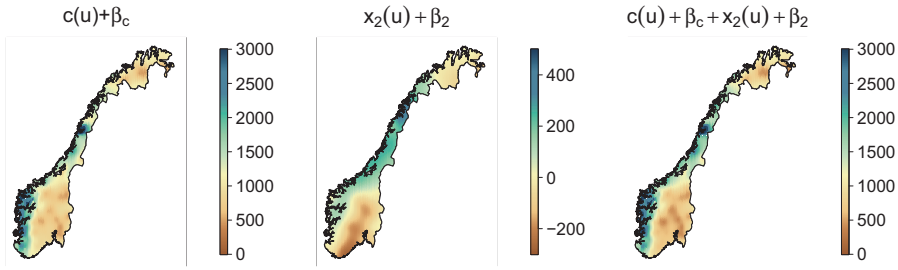
GRF $c(\boldsymbol{u})$ for our four datasets. These are shown in Table 2 and indicate how much of the spatial variability that is captured by the climatic GRF relative to the annual GRF. In particular, if σ_c dominates over σ_x, it suggests hydrological spatial stability.

Table 2: The posterior mode of the range parameters ρ_c and ρ_x and the marginal standard deviations σ_c and σ_x of the climatic and the annual GRFs $c(\boldsymbol{u})$ and $x_j(\boldsymbol{u})$ for the areal model (upper) and centroid model (lower). The posterior standard deviations of the parameters are shown in parenthesis as a measure of the uncertainty. The mode and standard deviations vary between the experiments and groups in the cross-validation, and the values given here are the mean over all folds and experiments (UG and PG). The spatial effect that dominates (annual or climatic) is marked in bold.

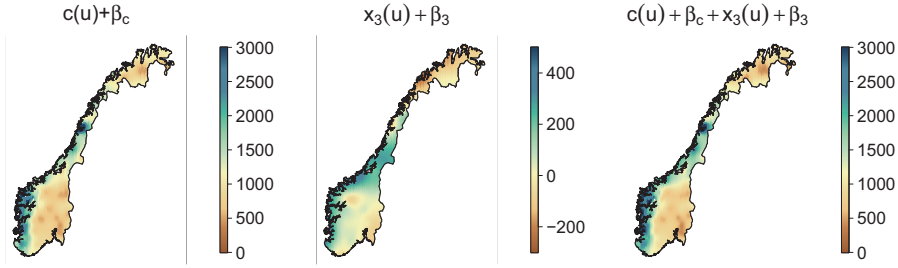
Areal model	ρ_c [km]	ρ_x [km]	σ_c [mm/year]	σ_x [mm/year]
Annual	58 (7)	476 (65)	**880** (56)	267 (23)
January	31 (7)	247 (22)	72 (6)	**83** (4)
April	77 (14)	239 (32)	**75** (6)	48 (3)
June	43 (5)	153 (22)	**181** (9)	75 (3)

Centr. model	ρ_c [km]	ρ_x [km]	σ_c [mm/year]	σ_x [mm/year]
Annual	89 (12)	659 (77)	**750** (57)	263 (22)
January	82 (15)	369 (44)	60 (6)	**88** (7)
April	118 (19)	375 (51)	**66** (4)	52 (5)
June	69 (9)	335 (47)	**161** (12)	71 (6)

The estimates in Table 2 show that the hydrological spatial stability is largest for June and for annual runoff, as expected from the time series in Figure 2b and Figure 4. Here, the posterior mode for σ_c is more than twice as large as the posterior mode for σ_x for both the areal and the centroid model. Furthermore, we see in Table 2 that the climatic range ρ_c is only around 12 % of the annual range ρ_x. In Figure 10 we have illustrated the spatial pattern these parameters give for annual predictions in 1997 and 1998 for the whole study area. We see that the annual runoff for 1997 and 1998 have the same spatial pattern, and that this spatial pattern mostly originates from $c(\boldsymbol{u})$, i.e. climatic conditions including catchment characteristics. The trend we see in Figure 10 can also be seen for the remaining eight years in the dataset (1996,1999-2005), as well as for June. A spatial pattern like this, with $\sigma_c \gg \sigma_x$ and $\rho_c < \rho_x$, suggests that the information gain from neighboring catchments further away is low for an



(a) Year 1997



(b) Year 1998

Figure 10: From left to right: The climatic part of the model (common for all years), the annual (year dependent) part of the model and the full model $q_j(\boldsymbol{u})$ for annual runoff in 1997 and 1998 [mm/year]. Note that the scales of the middle plots only cover 25 % of the scale of the other plots. We see that most of the spatial variability of annual runoff for 1997 and 1998 can be explained by climatic effects, and that the climatic range ρ_c is considerably smaller than the year specific range ρ_x. The results above are produced by the centroid model, and plots similar to these are behind all results presented for the areal and centroid model in this article.

ungauged catchment, and that the potential information stored in short records is high.

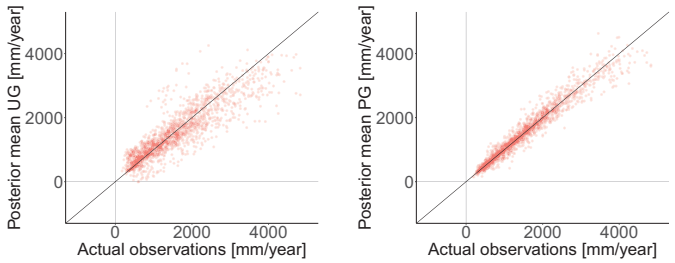
For January however the situation is different: The posterior mode of σ_x is larger than the posterior mode of σ_c for both the areal and the centroid model. The parameters show that for January, year-specific effects explain a larger part of the spatial variability. This can be due to a more unstable

hydrological setting with runoff driven by snow accumulation and snow melt. For April, we have that $\sigma_c > \sigma_x$, but σ_c is less dominant than for June and for the annual data.

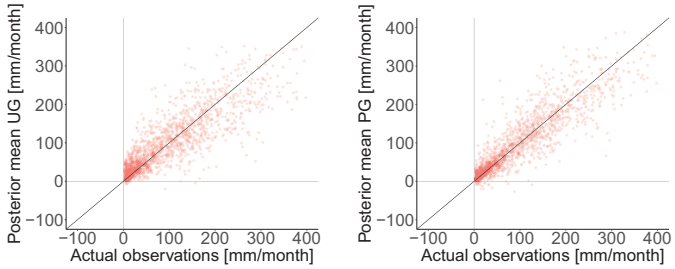
In the areal and centroid model, the inclusion of a short record changes the climatic spatial field $c(\boldsymbol{u})$, and hence the predictions can be considerably changed for the target catchment if the climatic effect is strong. The parameter values thus suggest that the gain of including short records is lower for April and January compared to the other two datasets. This is confirmed by comparing the RMSE and CRPS for the areal and the centroid model for the partially gauged case (PG), to the RMSE and CRPS obtained for the ungauged case (UG) in Table 1. For all datasets, the RMSE and CRPS for our two models are reduced for PG compared to UG, but the reduction is lower for January and April than for June and the annual data. For the annual predictions, the RMSE and CRPS are reduced by more than 50% when a short record of length one from the target catchment is included in the observation likelihood. The reduction for June is also remarkable (around 35-40 %), while the reduction for January and April is moderate (around 13-20 %). The results hold for both the areal and centroid model, but the areal model seems to be somewhat better than the centroid model in terms of exploiting short records of data from the target catchment. This is again related to the parameter estimates in Table 2, where we see that σ_c dominates more over σ_x in the areal model than in the centroid model.

Considering the results for Top-Kriging in Table 1, we only obtain a small reduction in the RMSE and CRPS for the partially gauged case (PG) compared to the ungauged case (UG). This is because Top-Kriging treats each year of data independently when considering infill of missing annual data. A reduction in RMSE and CRPS is only seen for the specific year with extra data. This is different from our framework where several years of data are modeled simultaneously. The evaluation scores in Table 1 and the boxplots in Figure 9 clearly show that our two suggested methods outperform Top-Kriging for the partially gauged case for annual predictions and monthly predictions in June, which were the two time-scales with most hydrological spatial stability ($\sigma_c \gg \sigma_x$). For January and April the three models are more similar in predictive performance.

For the PG case, we also compare the areal and the centroid model to simple linear regression. According to Table 1 and Figure 9 linear regression performs quite well for the annual data, which represent the most hydrologically spatially stable dataset. Linear regression actually provides the second lowest RMSE of all four methods for annual predictions. However, recall that a short record of length two from the target catchment is needed to use this method, while our areal model performs slightly better with a short record of length one (and observations from other neighboring catchments). For January, April and June, linear regression is outperformed by the three other methods in terms of RMSE and CRPS (Table 1).



(a) Annual predictions.



(b) Predictions for April

Figure 11: All observations for 1996-2005 compared to the corresponding predictions for the ungauged case (UG, left) and the partially gauged case (PG, right) for annual predictions (Figure 11a) and for April (Figure 11b). The predictions are performed by the areal model. The straight line represents a perfect correspondence between prediction and actual observation.

To illustrate the possible gain of including (very) short records of data

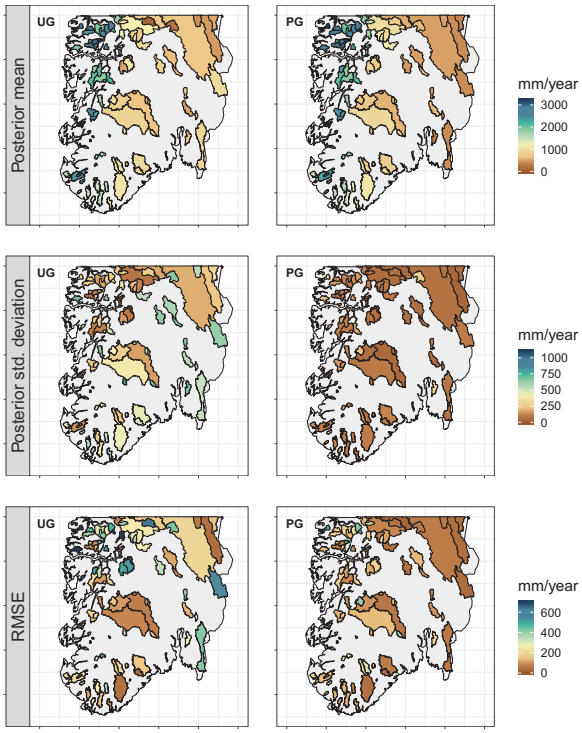


Figure 12: Average posterior mean $Q_j(\mathcal{A})$ (upper), average posterior standard deviation (middle) and RMSE (lower) for $j = 1, ..10$ for predictions of missing annual observations for the areal model for the ungauged case (UG, left) and the partially gauged case (PG, right).

from the target catchment, we present four scatter plots that compare the predicted values produced by the areal model to the actual observations of runoff (Figure 11). For the annual predictions in Figure 11a, the predictions for PG are considerably more concentrated around the straight line that indicates a perfect fit, than the predictions for UG. There are similar results for June, whereas the difference between the ungauged and partially gauged case is not that prominent for April (see Figure 11b) and January. Furthermore, the April scatter plots demonstrate that (very) short records

don't lead to a poorer predictive performance, even if April is a month driven by more unstable hydrological patterns. The predictions are simply not substantially affected by the new data points that are included in the likelihood, as we can see in Figure 11b. In our model, the risk of including very short records is low because climatic effects $c(\boldsymbol{u})$ are adjusted relative to year specific effects $x_j(\boldsymbol{u})$ by statistical inference. This way short records can safely be included in the modeling regardless of the underlying weather patterns and the degree of hydrological spatial stability.

In Figure 7 we saw that all three interpolation methods were able to reproduce the true spatial pattern of annual runoff when filling in missing annual values for ungauged catchments (UG). However, all three methods produced high RMSE values for some of the catchments. These were typically small catchments located on the western coast of Norway. Figure 12 shows the impact of including a short record of length one for these catchments. It compares the annual predictions from the ungauged case (UG) to the annual predictions from the partially gauged case (PG) for the areal model. We see a large reduction in the RMSE for many of the catchments, and a (realistic) reduction of the posterior standard deviation. We also see that a few of the catchments obtain a decrease in predictive performance when short records are included, but the overall tendency is clear: The gain of including short records for annual predictions in Norway is high, and the suggested framework is able to exploit this property.

6.2 Predictions of mean annual runoff (1981-2010)

So far, we have presented an evaluation of the framework's ability to fill in missing annual observations of runoff for a 10 year period (1996-2005). We now present the evaluation of the framework's ability to predict mean annual runoff for a 30 year period as a whole (1981-2010), as described in Section 5.2.

Figure 13 shows the RMSE and CRPS for the predictions of mean annual runoff for Top-Kriging, the areal and the centroid model as a function of record length (0, 1, 3, 5 and 10). The record length is the number of annual runoff observations available from the target catchment in the cross-validation. We find that Top-Kriging again performs best for the

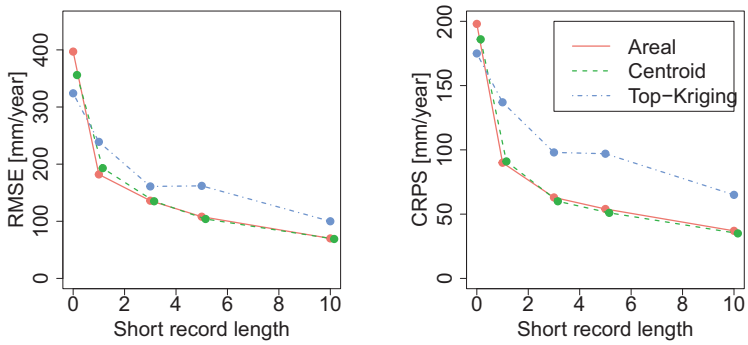


Figure 13: RMSE and CRPS as a function of record length (0, 1, 3, 5 and 10) for predictions of mean annual runoff for 1981-2010 for 83 fully gauged catchments in southern Norway.

ungauged case (short record length 0), while the centroid model performs slightly better than the areal model for ungauged target catchments. Furthermore, the RMSE and CRPS decrease with increasing record length for all three methods. However, Figure 13 shows that our areal and centroid models outperform Top-Kriging for record lengths larger than 0: The overall difference between our framework and Top-Kriging is around 30-60 mm/year in terms of RMSE, which is a considerable difference when the RMSE values are around 100-200 mm/year.

Furthermore, we notice the large increase in predictive performance when including a (very) short record of length one (PG1 in Figure 13). The reduction in RMSE and CRPS is 45-50 % from the UG to the PG1 case for the areal and centroid model. These results are thus comparable to the results we obtained for the experiments related to infill of missing annual values (Section 6.1).

To be able to compare our findings with other studies, we also included plots of the the absolute normalized error (ANE) and the squared correlation coefficient (r^2) for the experiments. These can be found in Figure 14 and 15, and are referred to in the discussion (Section 7.2). Also according to these scale independent evaluation criteria the overall results are that for ungauged catchments Top-Kriging performs best, while when there are

short records available, our framework performs better.

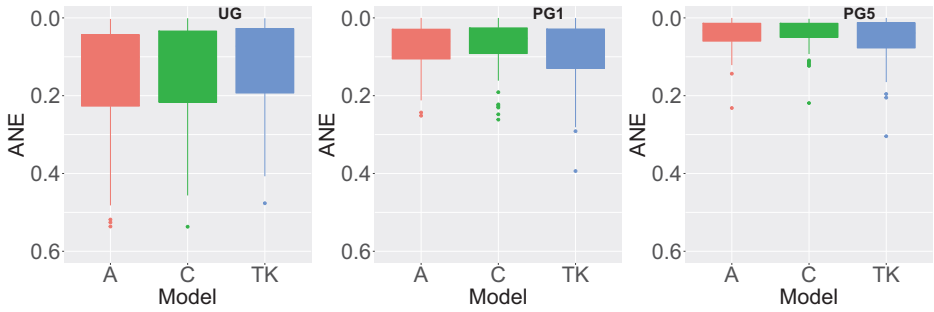


Figure 14: Absolute normalized error (ANE) for the areal model (A), centroid model (C) and Top-Kriging (TK) for predictions of mean annual runoff in ungauged catchments (UG, left) and in partially gauged catchments with short records of length one (PG1, middle) and length five (PG5 right).

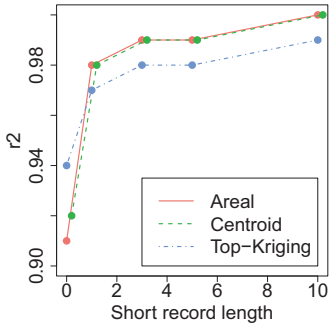


Figure 15: The squared correlation coefficient (r^2) for predictions of mean annual runoff for catchments in Norway with record length 0, 1, 3, 5 and 10.

7 Discussion

In this article we have presented a geostatistical framework particularly suitable for hydrological datasets that include short records of data. Here, we highlight four points for discussion: 1) the difference in performance across methods and study areas, 2) comparing the findings with other

studies, 3) shortcomings of the suggested framework and 4) suggested areas of use.

7.1 Difference in performance across methods and study areas

In our work, we evaluated two versions of our suggested framework by predicting annual runoff and mean annual runoff for Norway. The results showed that our areal referenced method and our point (or centroid) referenced method gave very similar results in terms of posterior mean (see e.g. Figure 7 and Figure 13). We did not find a trend describing when one of the methods performed better than the others. In prior to the analysis, we would expect the areal model to perform better than the centroid model for ungauged, nested catchments since the areal model takes the water balance and the nested structure of catchments into account. However, these properties did not have a notable impact on the predicted posterior mean runoff for this particular dataset. This is not an extraordinary result as similar results have been obtained by other studies that have compared Top-Kriging (areal referenced approach) to ordinary Kriging (point referenced approach): The point referenced approaches often perform similarly as the areal referenced approaches (Farmer, 2016; Skøien et al., 2014).

A possible explanation for the similar performance of the centroid and areal model in this study, is that the proportion of nested catchments in our datasets was relatively low: Only 30 % of the catchments in Figure 1 were nested, while the percentage of nested catchments was 53 % among the fully gauged catchments in Figure 5. Furthermore, most of the nested catchments only have one overlapping catchment. The water balance constraints of the areal model might be more important for datasets where there is a higher percentage of nested catchments in an area with high spatial variability. One example is shown in Roksvåg et al. (2020).

It is also possible that the water balance constraints of the areal model have some drawbacks. One example is if there is poor data quality for a subcatchment in the dataset. Then we impose an inaccurate, but relatively strict constraint on the runoff in this catchment's drainage area. This will have an impact on the predictions for all overlapping catchments, and how

the predicted runoff is distributed here. In this sense, the areal model is less flexible than the centroid model and requires better data quality.

The water balance constraints of the areal model also makes it computationally more expensive than the two other models. Top-Kriging used around 1 minute for the interpolation of mean annual runoff presented in Section 6.2 for one cross-validation fold. The centroid model used around 30-40 minutes for the interpolation, but provided results for both mean annual runoff and runoff for 30 individual years at the same time. Its run time is thus similar to the run time of Top-Kriging per year. The areal model on the other hand, used around 6-7 hours on the same computational server. Hence, from a practical point of view, the centroid model might be the most convenient version of our suggested framework for many applications. However, note that if the posterior uncertainty is important, the areal model gives a more realistic representation of uncertainty than the centroid model (see Figure 7). The centroid model also treats small and large catchments equally, which can be problematic for some applications and study areas.

When considering predictions for ungauged catchments, the results showed that Top-Kriging provided better results than our two suggested models. Figure 7 showed that the three methods failed and succeeded for many of the same catchments, but that our models failed slightly more than Top-Kriging on average. We also see an indication that our models fail more than Top-Kriging for ungauged catchments that are located further away from other catchments. See for example the catchment that is located south-east in Figure 7. For ungauged catchments located far away from other catchments (relative to the spatial range), the predicted value will go towards the intercept β_c for our two Bayesian models. For Top-Kriging, the predicted value will always be a weighted sum of the observations from the neighboring catchments. This can explain the difference in performance here. Apart from this, we don't find a pattern for which catchments Top-Kriging performs better (mean elevation, location and the magnitude of the observed value were investigated).

While Top-Kriging performed best for ungauged catchments, our framework outperformed Top-Kriging when there were some available data from the target catchment. This was the case both when predicting mean annual

runoff, and runoff for individual years. The results showed that the potential gain of including (very) short records in the modeling in Norway was large. An explanation is that the annual runoff in Norway is mainly controlled by orographic precipitation. Since the orographic precipitation is driven by topography and westerly winds are dominating, the precipitation patterns are repeated each year and we obtain hydrological spatial stability with $\sigma_c \gg \sigma_x$. The mountains in Norway also lead to rapid weather changes in space, here expressed through a low climatic spatial range ρ_c. Consequently, the information gain from neighboring catchments is often low for ungauged catchments, and information from the target catchment can be very valuable. It is also convenient that Norway has a humid climate where only around 10-20 % of the annual precipitation evaporates.

The evaluation study based on annual time series of monthly runoff gave us an indication of how the framework can be expected to behave for other climates and countries: For areas where the annual runoff is driven by unstable weather patterns and hydrological processes, short records can not be expected to contribute to as large improvements in the predictions as for the Norwegian annual data (see the predictions for April in Figure 11b). This might be the case for countries and areas where most of the runoff can be explained by convective precipitation, where the aridity index is large or for variables for which storage effects are significant. However, the monthly predictions for January and April also illustrated that we safely can include (very) short records in the model, even if year specific effects explain most of the spatial variability of runoff. By this we have demonstrated that our models represent a framework for safe use of short records regardless of record length and climate, and with the benefit that we don't need to consider the choice of donor catchment as in other comparable methods.

Norway is a country with a moderate gauging density. The framework has not been tested for a more dense gauging density. We suppose that there is less to gain from including short records if the gauging density is large relative to the spatial range: Here the information obtained from neighboring catchments could be sufficient. However, a high density of gauged catchments and a close distance to neighboring catchments does not always guarantee good predictability at an ungauged catchment (Patil and Stieglitz, 2011). It is for example often difficult to predict runoff in

ungauged catchments that are very small and/or located close to weather divides. We believe that for such catchments, our method for including short records can be useful regardless of gauging density (as long as the study area is characterized by repeated runoff patterns over time).

7.2 Comparing the findings of this study with other studies

There exist several other studies of mean annual runoff in the literature, and some of them are compared in terms of the absolute normalized error (ANE) in the chapter about annual runoff in Blöschl et al. (2013). According to Figure 5.27 in Blöschl et al. (2013), an ANE between 0.05 and 0.5 is a typical result for regions like Norway where the potential evapotranspiration is less than 40 % of the mean annual precipitation. Figure 14 showed that the median ANE obtained for our suggested models is around 0.12 for ungauged catchments, i.e. in the lower range of ANE values in Blöschl et al. (2013). When a short record og length one or five was available (PG1 and PG5), the median ANE was as low as 0.05 and 0.03 for our methods.

In Figure 5.30 in Blöschl et al. (2013) there is also a subplot showing the ANE for predictions of mean annual runoff for ungauged catchments in Austria. Here, geostatistical models (Top-Kriging) and process-based models (conceptual hydrological models) provided the best predictions according to the ANE, with a median ANE around 0.1. The results we obtain in Figure 14 for the ungauged catchments are thus comparable to the results from Austria. This is reasonable as the Austrian climate is humid, like the Norwegian, and the western part of the country is dominated by mountains (the Alps) and has similar climate characteristics as Norway.

Furthermore, Blöschl et al. (2013) reports an r^2 (squared correlation coefficient) between 0.60 and 0.99 for studies done by cross-validation of around 250 catchments, or for studies using models based on spatial proximity like our suggested framework (Figure 5.25 and Figure 5.26 in Blöschl et al. (2013)). The r^2 for our two models was shown in Figure 15, and we see that it lies between 0.91-0.99. This is in the higher range of values obtained by comparable studies.

7.3 Shortcomings

In this article, we proposed two models for runoff that are Gaussian. However, runoff is truncated at zero and typically not Gaussian distributed which we also can see from the histograms in Figure 1b and Figure 3. The consequence of the Gaussian assumptions is that there is nothing in the models that prevents them from predicting negative runoff. Negative values appear for both the areal and the centroid model due to the uncertainty given by σ_y, but this is also a problem for the Top-Kriging technique. Another source for negative values is that the climatic part of the model $(c(\boldsymbol{u}) + \beta_c)$ can be negative in some areas. This is a fully valid result because the other model components could still ensure positive predictions for most catchments and years. However, it can become a problem if we are unlucky and the year specific GRF doesn't make up for the negative climatic GRF for one specific year. To avoid negative values, it is possible to log transform the data before performing an analysis. However, this is only valid for the centroid model, as the log transform is not compatible with the linear aggregation performed by the areal model (Equation (11)).

In the areal model, negative values also appear as a consequence of requiring preservation of water balance. If there are inconsistent or poor data over nested catchments, negative runoff in parts of a catchment can be the only option to fulfill the water balance requirements. To avoid negative runoff it is important that the discretization of the study area is fine enough to capture rapid changes in runoff over nested catchments. Catchments that are significantly influenced by human activities should also be removed from the analysis as these can influence both the water balance and the significance of the climatic field $c(\boldsymbol{u})$ relative to the annual field $x_j(\boldsymbol{u})$.

In our study, some negative values were produced for the monthly predictions as we can see in Figure 11b. However, this is not common and happened for only 1.2 % of the predictions of missing monthly data, and for a few data points for the missing annual data for the areal and centroid model. For predictions of *mean* annual runoff, negative values almost never appear as such effects typically are averaged out. Note that unphysical results also appear for Top-Kriging and other interpolation methods,

either in terms of violating the water balance or in terms of negative values. These model weaknesses should be remarked such that the modeler is able to choose what is most important in a real modeling setting. In this case it is a choice between 1) avoiding negative values by log transforming the data before using Top-Kriging or the centroid model or 2) to impose water balance constraints through the areal model.

7.4 Suggested areas of use

Finally, we want to highlight what we think are the main areas of use for our suggested framework. First, our results showed that our main benefit compared to Top-Kriging was connected to exploiting short records from the *target* catchment. For this reason, we think that our method is suitable as a preprocessing method for making inference about the (mean) annual runoff in partially gauged catchments before doing a further analysis with other statistical tools or process-based models. One possible approach for runoff estimation could for example be a two step procedure where we (i) use the centroid or areal model as a record augmentation technique to predict runoff for the partially gauged catchments in the dataset, and (ii) use Top-Kriging to predict runoff in *ungauged* catchments. Here, the results from step (i) can be used as observed values in Top-Kriging together with data from fully gauged catchments. Differences in observation uncertainty between fully gauged- and partially gauged catchments should here be taken into account.

Secondly, we see that the parameter values of the suggested model provides interesting information about the study area. More specifically, if the marginal variance of the climatic GRF σ_c dominates over the marginal variance of the year specific GRF σ_x, it suggests that the spatial variability is stable over time, and that short records of runoff can have a large impact on the model, particularly if also $\rho_c < \rho_x$. This information can be used by decision-makers to e.g. motivate the installation of a new (possibly temporary) gauging station as this might improve the long-term estimates only a year after installation for this catchment. Likewise can the model and its parameters be used to assess whether a gauging station is redundant and can be shut down. However, to exactly quantify the importance of a

gauging station, all model variances (σ_x^2, σ_c^2, σ_β^2, σ_y^2) and ranges (ρ_x, ρ_c) must be taken into account, as well as the distances between the donor catchments and the target catchment. Computing this gain is outside the scope of this article, but an interesting topic for further research that is related to the field of decision theory and the *value of information* (Eidsvik et al., 2015).

8 Conclusions

We have presented a geostatistical framework for estimating runoff by modeling several years of runoff data simultaneously by using one (climatic) spatial field that is common for all years under study, and one (annual) spatial field that is year specific. By this, we obtain a framework that is particularly suitable for runoff interpolation when the available data originate from a mixture of gauged and partially gauged catchments, and that can be used to estimate runoff at ungauged and partially gauged locations. We evaluated the framework by 1) its ability to fill in missing values of annual runoff and 2) its ability to predict mean annual runoff for ungauged and partially gauged catchments. The case study from Norway showed that the suggested framework performs better than Top-Kriging for catchments that have short records of data, both for predictions of mean annual runoff and when filling in missing annual values. For totally ungauged catchments, Top-Kriging performed best. We also 3) demonstrated the potential value of including short records in the modeling and found that the value of (very) short records was high in Norway: An average reduction of 50 % in the RMSE was reported when a short record of length one was available from the target catchment, compared to when no annual observations were available. The reason for the large reduction is that the annual runoff in Norway is mainly driven by hydrological processes that are repeated each year. For such areas, our methodology has its main benefits, and we can use it as a tool for motivating the installation of new gauging stations: The new gauging stations might improve the long-term estimates at the target catchments only a year after installation. Furthermore, the results also show that the framework represents safe use of short records down to record lengths of one year, regardless of the underlying climatic

conditions in the area of interest.

Code availability

Example code for fitting the centroid model with example data is available on `https://github.com/tjroksva/RunoffInterpolation` (doi: 10.5281/zenodo.3630348). The remaining data are available upon request.

Acknowledgements

The project is funded by The Research Council of Norway, grant number: 250362. We would also like to thank Dr. Gregor Laaha, Dr. Jon Olav Skøien, Mr. Joris Beemster and one anonymous referee for in-depth reviews and valuable comments. The review process improved the quality of the paper.

References

K. Adamowski and C. Bocci. Geostatistical regional trend detection in river flow data. *Hydrological Processes*, 15(18):3331–3341, 2001. doi: 10.1002/hyp.1045.

H. Bakka, H. Rue, G.-A. Fuglstad, A. Riebler, D. Bolin, J. Illian, E. Krainski, D. Simpson, and F. Lindgren. Spatial modeling with r-inla: A review. *WIREs Computational Statistics*, 10(6), 2018. doi: 10.1002/wics.1443.

S. Banerjee, A.E. Gelfand, and B.P Carlin. *Hierarchical Modeling and Analysis for Spatial Data*, volume 101 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 2004.

G. Blöschl, M. Sivapalan, T. Wagener, A. Viglione, and H. Savenije. *Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales.* Camebridge University press, 2013.

S. Brenner and L. Scott. *The Mathematical Theory of Finite Element Methods, 3rd Edition. Vol. 15 of Texts in Applied Mathematics*. Springer, 2008.

G. Casella and R.L. Berger. *Statistical Inference*. Duxbury Press Belmont, 1990.

N. Cressie. *Statistics for spatial data*. J. Wiley & Sons, 1993.

J. Eidsvik, A. O. Finley, S. Banerjee, and H. Rue. Approximate bayesian inference for large spatial datasets using predictive process models. *Computational Statistics & Data Analysis*, 56(6):1362 – 1380, 2012. doi: 10.1016/j.csda.2011.10.022.

J. Eidsvik, T. Mukerji, and D. Bhattacharjya. *Value of Information in the Earth Sciences: Integrating Spatial Modeling and Decision Analysis*. Cambridge University Press, 2015. doi: 10.1017/CBO9781139628785.

K. Engeland and H Hisdal. A comparison of low flow estimates in ungauged catchments using regional regression and the hbv-model. *Water Resources Management*, 23: 2567–2586, 09 2009. doi: 10.1007/s11269-008-9397-7.

W. Farmer. Ordinary kriging as a tool to estimate historical daily streamflow records. *Hydrology and Earth System Sciences Discussions*, 2016:1–23, 01 2016. doi: 10.5194/hess-2015-536.

E. Ferkingstad and H. Rue. Improving the inla approach for approximate bayesian inference for latent gaussian models. *Electronic Journal of Statistics*, 9, 12 2015. doi: 10.1214/15-EJS1092.

M.B. Fiering. Use of correlation to improve estimates of the mean and variance. *USGS Publications Warehouse*, 1963.

Y. Fong, H. Rue, and J. Wakefield. Bayesian inference for generalized linear mixed models. *Biostatistics (Oxford, England)*, 11:397–412, 12 2009. doi: 10.1093/biostatistics/kxp053.

G-A. Fuglstad, D. Simpson, F. Lindgren, and H. Rue. Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association*, 114(525):445–452, 2019. doi: 10.1080/01621459.2017.1415907.

E. J. Førland. *Nedbørens høydeavhengighet*. Klima, 1979.

D. Gamerman and H. F Lopes. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC, 2006.

T. Gneiting and A. E. Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102:359–378, 2007. doi: 10.1198/016214506000001437.

L. Gottschalk. Correlation and covariance of runoff. *Stochastic Hydrology and Hydraulics*, 7:85–101, 1993. doi: 10.1007/BF01581418.

L. Gottschalk, J. L. Jensen, D. Lundquist, R. Solantie, and A. Tollan. Hydrologic regions in the nordic countries. *Hydrology Research*, page 273–286, 1979.

G. Guillot, R. Vitalis, A. le Rouzic, and M. Gautier. Detecting correlation between allele frequencies and environmental variables as a signature of selection. a fast computational approach for genome-wide studies. *Spatial Statistics*, 8:145 – 155, 2014. ISSN 2211-6753. doi: 10.1016/j.spasta.2013.08.001.

P. Guttorp and T. Gneiting. Studies in the history of probability and statistics XLIX on the matérn correlation family. *Biometrika*, 93(4):989–995, 2006. doi: 10.1093/biomet/93.4.989.

R. M. Hirsch. A comparison of four record extension techniques. *Water Resour. Res.*, 1982.

J. Huang, B. Malone, B. Minasny, A. Mcbratney, and J. Triantafilis. Evaluating a bayesian modelling approach (inla-spde) for environmental mapping. *Science of The Total Environment*, 609:621–632, 12 2017. doi: 10.1016/j.scitotenv.2017.07.201.

R. Ingebrigtsen, F. Lindgren, and I. Steinsland. Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, 8:20 – 38, 2014. doi: 10.1016/j.spasta.2013.06.002.

R. Ingebrigtsen, F. Lindgren, I. Steinsland, and S. Martino. Estimation of a non-stationary model for annual precipitation in southern Norway using replicates of the spatial field. *Spatial Statistics*, 14:338 – 364, 2015. doi: 10.1016/j.spasta.2015.07.003.

D. Khan and M. Warner. A bayesian spatial and temporal modeling approach to mapping geographic variation in mortality rates for subnational areas with r-inla. *Journal of data science: JDS*, 18:147–182, 01 2018.

G. Laaha and G. Blöschl. Low flow estimates from short stream flow records — a comparison of methods. *Journal of Hydrology*, 306(1):264 – 286, 2005. doi: 10.1016/j.jhydrol.2004.09.012.

F. Lindgren, H. Rue, and J. Lindström. 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)*, 73:423–498, 2011. doi: 10.1111/j.1467-9868.2011.00777.x.

S. Martino, R. Akerkar, and H. Rue. Approximate bayesian inference for survival models. *Scandinavian Journal of Statistics*, 38:514 – 528, 09 2011. doi: 10.1111/j.1467-9469.2010.00715.x.

N. C. Matalas and B. Jacobs. A correlation procedure for augmenting hydrologic data. *U.S. Geol. Surv. Prof. Pap.*, 434-E(E1-E7), 1964.

R. Merz and G. Blöschl. Flood frequency regionalisation - spatial proximity vs. catchment attributes. *Journal of Hydrology*, 302:283 – 306, 2005. doi: 10.1016/j.jhydrol.2004.07.018.

P. Moraga, S. M. Cramb, K. L. Mengersen, and M. Pagano. A geostatistical model for combined analysis of point-level and area-level data using INLA and SPDE. *Spatial Statistics*, 21:27 – 41, 2017. doi: 10.1016/j.spasta.2017.04.006.

T. Opitz, R. Huser, H. Bakka, and H. Rue. Inla goes extreme: Bayesian tail regression

for the estimation of high spatio-temporal quantiles. *Extremes*, 21, 02 2018. doi: 10.1007/s10687-018-0324-x.

S. Patil and M. Stieglitz. Controls on hydrologic similarity: role of nearby gauged catchments for prediction at an ungauged catchment. *Hydrology and Earth System Sciences Discussions*, 8, 10 2011. doi: 10.5194/hessd-8-9323-2011.

A. Petersen-Øverleir. Accounting for heteroscedasticity in rating curve estimates. *Journal of Hydrology*, 292:173–181, 06 2004. doi: 10.1016/j.jhydrol.2003.12.024.

T. Roksvåg, I. Steinsland, and K. Engeland. A geostatistical two field model that combines point observations and nested areal observations, and quantifies long-term spatial variability - a case study of annual runoff predictions in the voss area. *arXiv:1904.02519v2*, 2020.

H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 2005.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71:319–392, 2009. doi: 10.1111/j.1467-9868.2008.00700.x.

E. Sauquet, L. Gottschalk, and E. Lebois. Mapping average annual runoff: A hierarchical approach applying a stochastic interpolation scheme. *Hydrological Sciences Journal*, 45(6):799–815, 2000. doi: 10.1080/02626660009492385.

S. Siegel. *Non-parametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

D. Simpson, H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, 32:1–28, 2017. doi: 10.1214/16-STS576.

J. O. Skøien, R. Merz, and G. Blöschl. Top-kriging - geostatistics on stream networks. *Hydrology and Earth System Sciences Discussions*, 10:277–287, 2006. doi: 10.5194/hess-10-277-2006.

J. O. Skøien and G. Blöschl. Spatiotemporal topological kriging of runoff time series. *Water Resources Research*, 43(9), 2007. doi: 10.1029/2006WR005760.

J. O. Skøien, G. Blöschl, and A. W. Western. Characteristic space scales and timescales in hydrology. *Water Resources Research*, 39, 2003. doi: 10.1029/2002WR001736.

J.O. Skøien, G. Blöschl, G. Laaha, E. Pebesma, J. Parajka, and A. Viglione. rtop: An r package for interpolation of data with a variable spatial support, with an example from river networks. *Computers & Geosciences*, 67:180 – 190, 2014. ISSN 0098-3004. doi: https://doi.org/10.1016/j.cageo.2014.02.009.

A. Stohl, C. Forster, and H. Sodemann. Remote sources of water vapor forming precipitation on the Norwegian west coast at 60 °N - a tale of hurricanes and an atmospheric river. *Journal of Geophysical Research: Atmospheres*, 113(D5), 2008. doi: 10.1029/2007JD009006.

N. Sælthun, O. Tveito, T. Bøsnes, and L. Roald. Regional flomfrekvensanalyse for norske vassdrag. *Tech. Rep. Oslo: NVE*, 1997.

A. Viglione, J. Parajka, M. Rogger, J L. Salinas, G. Laaha, M. Sivapalan, and G Blöschl. Comparative assessment of predictions in ungauged basins - Part 3: Runoff signatures in Austria. *Hydrology and Earth System Sciences Discussions*, 10:449–485, 2013. doi: 10.5194/hessd-10-449-2013.

Richard M. Vogel and Jery R. Stedinger. Minimum variance streamflow record augmentation procedures. *Water Resources Research*, 21(5):715–723, 1985. doi: 10.1029/WR021i005p00715.

Y. Yuan, F. Bachl, F. Lindgren, D. Borchers, J. Illian, S. Buckland, H. Rue, and T. Gerrodette. Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *The Annals of Applied Statistics*, 11:2270–2297, 12 2017. doi: 10.1214/17-AOAS1078.

Paper C

A geostatistical spatially varying coefficient model for mean annual runoff that incorporates process-based simulations and short records

Thea Roksvåg, Ingelin Steinsland and Kolbjørn Engeland

Technical note

An overview of a geostatistical two field model for interpolation of environmental variables

---

Thea Roksvåg

Technical note

An overview of a geostatistical two field model for interpolation of environmental variables

Thea Roksvåg

1 Introduction

Environmental variables like temperature, precipitation and runoff are in many areas of the world driven by weather patterns that are repeated over time. These weather patterns are typically caused by wind patterns, seasonality of temperature and the topography in the study area. In this note, we highlight a specific spatial model that is particularly suitable for modeling (environmental) variables of this type. It is a Bayesian hierarchical model for modeling spatio-temporal data that consists of two spatial fields: One of the fields is constant over time and captures the long-term spatial variability or the climate in the study area. The other spatial field is connected to a specific time unit, typically a year, and is used to model the spatial variability that cannot be explained by long-term effects. Because of the geostatistical model's two characteristic spatial fields, we refer to it as *the two field model*.

The two field model was developed and presented in Roksvåg et al. (2020a,b) where it was successfully used for interpolation of Norwegian annual runoff, a hydrological variable that tends to follow repeated spatial patterns over time due to orographic precipitation and the topography in the study area. During our work with these papers, we did some experiments that showed

that the two field model can be useful also outside the field of runoff interpolation: Environmental variables such as annual and monthly precipitation or temperature follow similar spatial patterns as the Norwegian annual runoff. For this type of data structures, the two field model has some beneficial properties that other spatial and spatio-temporal models don't have. The aim of this note is to highlight some of these properties and explore the two field model further. By this we answer some of the questions that were considered outside the scope of Roksvåg et al. (2020a,b), where the main focus was on performing runoff predictions.

One of the central questions is how the two field model performs compared to other statistical models for spatial and spatio-temporal data. In the note we investigate this through an illustrative case example where we highlight model differences between the two field model, a standard one field spatial model and a comparable spatio-temporal model. We describe for what kinds of data structures the two field model has its benefits, and when the two other models can be used instead. To explore the models we also use simulated data such that we can control the model parameters and the observation set-up. In Roksvåg et al. (2020a,b) the two field model was only compared to state-of-the-art hydrological interpolation approaches and not to the approaches we discuss in this note. Hence, the experiments presented here represent new contributions.

Next, we investigate if the two field model is well behaved in terms of parameter identifiability. This was another unanswered question from Roksvåg et al. (2020a,b). To investigate the parameter identifiability, we present a simulation study. We are particularly interested in confirming that it is possible to separate the two spatial fields from each other in the two field model. This is important as the parameters of the two spatial fields turn out to be closely linked to the predictability of the target variable, as we soon will see.

The major objective of this note is to give an overview of the two field model and describe for what data and applications it can be useful. Based on our experiments and the work in Roksvåg et al. (2020a,b), we hence suggest

some areas of use. We also present a collection of new case examples. Finally, we provide a tutorial for implementing the two field model in `r-inla`. This is a commonly used `R` package for performing approximate inference on Bayesian hierarchical models that are computationally demanding (Rue et al., 2009).

Before presenting the experiments and case examples outlined above, we specify the two field model, the spatial reference model and the spatio-temporal reference model. In the following we assume that the reader has knowledge about Gaussian random fields and Bayesian hierarchical models. We refer to Roksvåg et al. (2020a,b) for more methodological background and detailed reference lists.

2 Models under discussion

In this section we present the two field model from Roksvåg et al. (2020a,b) and two reference models. The two field model is applicable for spatio-temporal data. Hence, we compare it to a spatio-temporal model. However, often are spatio-temporal data modeled by simply interpolating the data for different time units independently, without considering temporal dependency. This is done either because the temporal dependency is difficult to see from the data or because most spatio-temporal models are computationally demanding. Motivated by this, we also compare the two field model to a standard spatial one field model. The comparison is particularly relevant because the environmental data we discuss in this note are associated with a temporal dependency structure that is less intuitive from studying the data.

We operate in a Bayesian hierarchical framework and the three models are specified such that they are suitable for approximate inference with INLA and SPDE (Rue et al., 2009; Lindgren et al., 2011). See Section 2.4 for more.

2.1 The two field model

The model we discuss in this note, is the two field spatial model from Roksvåg et al. (2020a,b). It is a Bayesian geostatistical model with two spatial fields, where one of the spatial fields represents the climate of the variable of interest, while the other spatial field represents the discrepancy from the climate for a specific time unit. To make the model more intuitive and the presentation simpler, we use *annual precipitation* as an example response variable throughout the text.

Let the spatial process $\{\eta_j(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ denote the true annual precipitation at point location $\boldsymbol{s}$ in the spatial domain $\mathcal{D} \in \mathcal{R}^2$ in year j. The true annual precipitation at location $\boldsymbol{s}$ in year j is modeled as

$$
\begin{aligned}
&\eta_j(\boldsymbol{s}) = \beta_c + c(\boldsymbol{s}) + \beta_j + x_j(\boldsymbol{s}) && (1)\\
&\pi(c(\boldsymbol{s})|\rho_c, \sigma_c) \sim \text{GRF}(0, \boldsymbol{\Sigma}(\rho_c, \sigma_c)) && \pi(\beta_c) \sim \mathcal{N}(\cdot, \cdot)\\
&\pi(x_j(\boldsymbol{s})|\rho_x, \sigma_x) \sim \text{GRF}(0, \boldsymbol{\Sigma}(\rho_x, \sigma_x)) && \pi(\beta_j|\tau_\beta^{-1}) \sim \mathcal{N}(\cdot, \tau_\beta^{-1}).
\end{aligned}
$$

Here, the parameter β_c is an intercept common for all years $j = 1, ..., r$, while $c(\boldsymbol{s})$ is a spatial effect common for all years. These two model components represent the part of the spatial process that is caused by long-term patterns, or the climate in the study area. Further, we include a year specific intercept β_j and a year specific spatial effect $x_j(\boldsymbol{s})$ to model the annual discrepancy from the climate in year j. We denote $c(\boldsymbol{s})$ the *climatic spatial field* and $x_j(\boldsymbol{s})$ the *replicated spatial field*. Both spatial fields are modeled as stationary Gaussian random fields (GRFs) given the model parameters, with zero mean and covariance matrix $\boldsymbol{\Sigma}$. The covariance matrices are constructed by a covariance function; $c(\boldsymbol{s})$ with range parameter ρ_c and marginal variance σ_c^2, and $x_j(\boldsymbol{s})$ with range parameter ρ_x and marginal variance σ_x^2. Furthermore, the spatial fields $x_j(\boldsymbol{s})$ for $j, ..., r$, are assumed to be independent realizations, or replicates of the underlying GRF. This implies that the covariance between $x_j(\boldsymbol{s}_i)$ and $x_v(\boldsymbol{s}_k)$, conditional on the model parameters for any two locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_k$, is zero for $v \neq j$. Similarly, are the year specific intercepts $\{\beta_j\}$ treated as replicates over time. More

specifically are they assumed to be independent and identically distributed as $\mathcal{N}(0, \tau_\beta^{-1})$ given the parameter τ_β, with $\beta_1, ..., \beta_r$ being independent realizations of this Gaussian distribution.

To gain knowledge about the true annual precipitation $\eta_j(\boldsymbol{s})$, precipitation is observed at n locations $\boldsymbol{s}_i \in \mathcal{D}$ for $i = 1, .., n$ over r years j=1,..r. The observations are associated with measurement uncertainty, and we model the observed precipitation as

$$y_{ij} = \eta_j(\boldsymbol{s}_i) + \epsilon_{ij}, \tag{2}$$

where y_{ij} is the observation at location $\boldsymbol{s}_i$ for year j, $\eta_j(\boldsymbol{s}_i)$ is the true annual precipitation, and ϵ_{ij} are error terms that are assumed to be independent and identically distributed with prior $\mathcal{N}(0, \tau_\epsilon^{-1})$. From this, the observation likelihood of the suggested model can be written as

$$\begin{aligned}\pi(\boldsymbol{y}|\boldsymbol{x^*}, \tau_y^{-1}) \sim \prod_{i=1}^{n}\prod_{j=1}^{r}(I\{\text{Observation } \ y_{ij} \text{ is available}\} \cdot \mathcal{N}(\eta_j(\boldsymbol{s}_i), \tau_\epsilon^{-1}) \\ + 1 \cdot I\{\text{Observation } \ y_{ij} \ \text{ is missing}\}),\end{aligned} \tag{3}$$

where $\boldsymbol{y}$ holds the observed annual precipitation y_{ij} for all observation locations $i = 1, ..., n$ and years $j = 1, ..., r$, $\eta_j(\boldsymbol{s}_i)$ is the underlying true annual precipitation from Equation (1) and $\boldsymbol{x^*}$ is a vector containing all latent variables, i.e the intercepts β_c, β_j and the GRFs $c(\boldsymbol{s}.)$ and $x_j(\boldsymbol{s}.)$ for all observation locations and years. Furthermore, we include an indicator function $I(\cdot)$ that is equal to 1 if its argument is true, and 0 otherwise. This allows for missing data: For some of the n locations there might be less than r observations available.

We operate in a Bayesian framework and specify the suggested model as a Bayesian hierarchical model with six so-called hyperparameters $\boldsymbol{\theta} = (\tau_\epsilon, \rho_x, \sigma_x, \rho_c, \sigma_c, \tau_\beta)$. The hierarchical model has three stages that are built by the data model defined in Equation (3), the process model defined in Equation (1) and the prior distributions of the hyperparameters

$\pi(\boldsymbol{\theta})$. We will not discuss prior specification for hyperparameters in this note. Also mark that the intercept β_0 is considered as a part of the process model, and not as a hyperparameter, as this fits into the INLA framework. For an introduction to hierarchical modeling and spatial statistics we refer to Banerjee et al. (2004) and Cressie (1993); Gelfand et al. (2010); Lieshout (2019) respectively.

2.2 Spatial reference model: A one field model

As a reference model for the two field model, we use a simpler spatial model that we denote *the one field model*. To specify the one field model, we use the same framework as in Section 2.1, except that we omit the climatic spatial field in Equation (1). Hence, in this model the true annual precipitation in year j is given by

$$\eta_j(\boldsymbol{s}) = \beta_c + \beta_j + x_j(\boldsymbol{s}), \tag{4}$$

and the prior distributions of β_c, β_j and $x_j(\boldsymbol{s})$ are as in Equation (1). The observation likelihood is also specified similarly as in Equation (3). The one field model treats each time unit j of data independently, except that x_j and β_j are modeled as replicates of underlying processes that are driven by the same parameters for year $j = 1, ..r$. This is different from the two field model that in addition accounts for time dependency through the climatic spatial field $c(\boldsymbol{s})$.

2.3 Spatio-temporal reference model

The two field model is able to borrow strength in time and space, and it therefore makes sense to also compare it to a spatio-temporal model. A comparable spatio-temoral model is defined here.

Let the spatio-temporal process $\{\eta(\boldsymbol{s}, j), (\boldsymbol{s}, j) \in \mathcal{D} \in \mathcal{R}^2 \times \mathcal{R}\}$ denote the annual precipitation at a point location $\boldsymbol{s} \in \mathcal{R}^2$ in year $j \in \mathcal{R}$. For simplicity

we denote the process $\eta_j(\boldsymbol{s})$, such that the notation is similar to the notation used in the previous two models. We define a comparable spatio-temporal model for the true annual precipitation at location $\boldsymbol{s}$ in year j as follows:

$$\begin{aligned} \eta_j(\boldsymbol{s}) &= \beta_c + \beta_j + w_j(\boldsymbol{s}) \\ w_j(\boldsymbol{s}) &= a w_{j-1}(\boldsymbol{s}) + x_j(\boldsymbol{s}); \quad j = 2, .., r. \end{aligned} \tag{5}$$

Here, β_c is an intercept common for all years while the β_j's for $j = 1, .., r$ are year specific intercepts. These two model components are defined as before in Section 2.1. The component $w_j(\boldsymbol{s})$ represents the spatio-temporal part of the model and changes in time according to a first-order autoregressive process that depends on the previous time step $w_{j-1}(\boldsymbol{s})$ through a parameter a. Here, $|a| < 1$, and the first time step is distributed as $w_1(\boldsymbol{s}) \sim \mathcal{N}(0, \frac{\sigma^2}{1-a^2})$. The component $x_j(\boldsymbol{s})$ in Equation (5) is a Gaussian field that follows the following covariance model:

$$\text{Cov}(x_j(\boldsymbol{s}_i), x_v(\boldsymbol{s}_k)) = \begin{cases} 0 & \text{if } j \neq v \\ \text{Cov}(x_j(\boldsymbol{s}_i), x_j(\boldsymbol{s}_k)) & \text{if } j = v. \end{cases} \tag{6}$$

Over different time units, the covariance of the GRF is 0 ($j \neq v$). Hence, the spatio-temporal covariance model is separable and can be written as the product of a purely spatial and a purely temporal covariance function. The covariance $\text{Cov}(x_j(\boldsymbol{s}_i), x_j(\boldsymbol{s}_k))$ between two locations $\boldsymbol{s}_i$ and $\boldsymbol{s}_k$ within a specific year $j = v$ is given by element (i, k) in a covariance matrix $\boldsymbol{\Sigma}(\rho_x, \sigma_x^2)$. To make the models comparable, we use the same covariance model here as we use for the GRF $x_j(\boldsymbol{s})$ in the models defined in Section 2.1 and Section 2.2.

The spatio-temporal model is formulated as a Bayesian hierarchical model similarly as in Section 2.1, except that we use the process model from Equation (5) instead of the process model from Equation (1). We refer to Blangiardo and Cameletti (2015) for more on spatio-temporal models.

2.4 Approximate inference

The three models formulated in the previous subsections are three-staged Bayesian hierarchical models. Making inference on them implies assessing the posterior distributions of their model parameters and their spatial field(s) at all target locations, given data. This is computationally demanding. To solve the computational challenges, we fit the models by using integrated nested Laplace approximations (INLA). The INLA methodology was suggested by Rue et al. (2009) and is a tool for making approximate Bayesian inference on latent Gaussian models (LGMs), i.e. hierarchical models where the process model consists of components that are given Gaussian priors. As the latent field $\boldsymbol{x}^*$ in Equation (3) holds the Gaussian random fields $x_j(\boldsymbol{s})$ and $c(\boldsymbol{s})$ and intercepts β_j and β_c with Gaussian priors, the above models are LGMs suitable for INLA.

In order to achieve fast predictions with INLA, the precision matrix (i.e. the inverse covariance matrix) of the latent field $\boldsymbol{x}^*$ should be sparse with many zero elements. This is typically not the case when working with Gaussian random fields. To solve this problem, we use the SPDE approach to spatial statistics. The SPDE approach was suggested by Lindgren et al. (2011). It can be used to approximate Gaussian random fields (GRFs) with Matérn covariance functions by Gaussian Markov random fields (GMRFs). This is achieved by solving a stochastic partial differential equation (SPDE) numerically. The SPDE approach comes with computational benefits, as GMRFs typically have precision matrices with more zero elements: Efficient algorithms for matrix operations on such matrices exist (Rue and Held, 2005). We refer to Rue et al. (2009) and Lindgren et al. (2011) for details on the INLA and SPDE approaches.

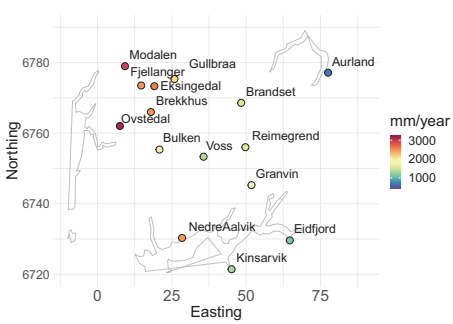
3 Illustrative example: Interpolation of annual precipitation with a dataset with missing values

In this section, we illustrate some of the differences between the two field model, the one field model and the spatio-temporal reference model by a motivational example. For this purpose, we use real annual precipitation data from Voss in western Norway and perform predictions in time and space. In addition we comment the two field model's parameter estimates.

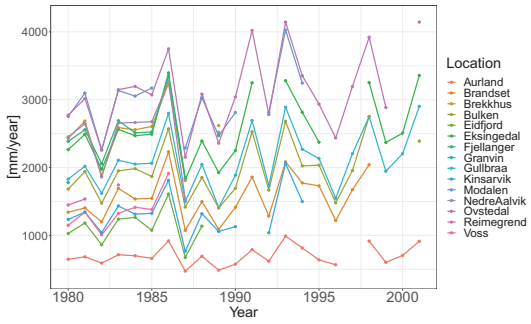
3.1 Data and set-up

To illustrate the differences between the three models from Section 2, we use a small dataset consisting of observations of annual precipitation from the Voss area in Norway from 1980-2001, i.e $\eta_j(\boldsymbol{s})$ is the annual precipitation in year $j = 1, ...22$. The observations were downloaded from `eklima.no`, a web page that gives free access to the climate database of the Norwegian Meteorological Institute. In the selected dataset, there are observations of annual precipitation from $n = 15$ point locations for $r = 22$ years, but none of the precipitation stations have observations for all years in the time period of interest. This is a typical situation for environmental variables. The locations of the precipitation gauges are shown in Figure 1a with corresponding time series in Figure 1b. We also include the full dataset in matrix form in Figure 1c. The goal in this example is to perform spatial interpolation of precipitation and make predictions for locations that lack observations, i.e. to fill in the `NA`'s in the data matrix in Figure 1c. We also use our models to make predictions for unobserved future years. Mark that the same dataset were used in Roksvåg et al. (2020a) to construct point referenced runoff observations, but the precipitation observations were not analyzed independently as in this example.

In the motivational example we use a stationary Matérn covariance function to construct the covariance matrices $\boldsymbol{\Sigma}(\rho_x, \sigma_x)$ and $\boldsymbol{\Sigma}(\rho_c, \sigma_c)$ for our models. Matérn covariance matrices make the models suitable for approx-



(a) Mean annual precipitation 1980-2001 around Voss.



(b) Time series.

.	Brandset	Modalen	Gullbraa	Eksingedal	Fjellanger	Reimegrend	Voss	Bulken	Brekkhus	Ovstedal	NedreAalvik	Granvin	Eidfjord	Kinsarvik	Aurland
1980	1342	2754	1829	2264	2387	1448	1149	1684	2432	2773	2452	1778	1029	1236	648
1981	1404	3098	2019	2491	2556	1536	1343	1940	2686	3017	2645	NA	1183	1342	684
1982	1199	2256	1620	1977	2057	NA	1008	1474	1882	2262	1862	NA	862	1041	592
1983	1693	3136	2107	2562	2691	1744	1324	1951	2589	3148	2655	NA	1239	1433	714
1984	1536	3054	2051	2470	2513	NA	1414	1984	2557	3197	2664	NA	1266	1312	700
1985	1546	3172	2064	2492	2524	NA	1383	1868	2606	3073	2676	NA	1076	1325	661
1986	2232	NA	2799	3395	3351	NA	1913	2571	3240	3750	3264	NA	1610	1811	920
1987	1073	2288	1499	1822	1849	NA	NA	1417	1811	2154	1535	NA	674	765	473
1988	1498	3027	2044	2391	NA	NA	NA	1852	NA	3084	NA	NA	1137	1319	693
1989	1089	2468	1411	1926	2485	NA	NA	1404	2618	2359	2520	NA	NA	1058	487
1990	1416	2812	1888	2251	NA	NA	NA	1693	NA	3041	NA	NA	NA	1129	576
1991	1858	NA	2693	3249	NA	NA	NA	2528	NA	4022	NA	NA	NA	NA	789
1992	1285	2816	1733	NA	NA	NA	NA	1668	NA	2781	NA	NA	NA	1038	620
1993	2084	4027	2890	3280	NA	NA	NA	2681	NA	4144	NA	NA	NA	2060	990
1994	1773	3246	2270	2815	NA	NA	NA	2024	NA	3353	NA	NA	NA	1500	815
1995	1730	NA	2133	2374	NA	NA	NA	2035	NA	2935	NA	NA	NA	NA	638
1996	1219	NA	1547	NA	NA	NA	NA	1481	NA	2436	NA	NA	NA	NA	567
1997	1674	NA	2205	NA	NA	NA	NA	1956	NA	3195	NA	NA	NA	NA	NA
1998	2041	NA	2750	3251	NA	NA	NA	2753	NA	3921	NA	NA	NA	NA	916
1999	NA	NA	1945	2369	NA	NA	NA	NA	NA	2885	NA	NA	NA	NA	602
2000	NA	NA	2204	2508	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	705
2001	NA	NA	2902	3359	NA	NA	NA	2390	NA	4144	NA	NA	NA	NA	914

(c) Full dataset.

Figure 1: Precipitation data from Norway [mm/year] for the illustrative example. The coordinate system is utm33N given in km.

imate inference with the SPDE approach to spatial modeling. Prior specification is not discussed in this note, but for completeness and further reference we specify the prior distributions we use for the analysis here. For most of the parameters we use PC priors (Simpson et al., 2017). For the precision of the year specific intercept τ_β we use a PC prior specified as $\text{Prob}(1/\sqrt{\tau_\beta} > 1000 \text{ mm/year}) = 0.2$ while for the precision of the measurement uncertainty τ_ϵ we use a PC prior specified as $\text{Prob}(1/\sqrt{\tau_\epsilon} > 1000 \text{ mm/year}) = 0.1$. For the parameters of the two Matérn fields $c(\boldsymbol{s})$

and $x_j(\boldsymbol{s})$ we use the joint PC prior for the spatial range and the marginal variance suggested by Fuglstad et al. (2019). These priors are specified as $\text{Prob}(\rho_x < 10 \text{ km }) = 0.1$, $\text{Prob}(\sigma_x < 2000 \text{ mm/year }) = 0.1$, $\text{Prob}(\rho_c < 10 \text{ km }) = 0.1$ and $\text{Prob}(\sigma_c < 2000 \text{ mm/year }) = 0.1$, i.e. we use the same priors for the two spatial fields. For the climatic intercept β_c we use a normal prior $\mathcal{N}(0, 10^7 \text{ mm/year})$. We use the same priors for all models in this illustrative example: The two field model, the one field model and the spatio-temporal model. For the a parameter of the spatio-temporal model we use the default `R-INLA` prior, $\text{loggamma}(1, 5 \cdot 10^{-5})$.

3.2 Comparing the prediction results across models

The one field model, the two field model and the spatio-temporal model were fitted to the dataset in Figure 1. Figure 2 shows the average posterior mean precipitation for 1981-2010. Here, the two field model and the spatio-temporal model look similar, although the spatio-temporal model seems to give a larger spatial range. The one field model however, looks more different from the two other models, particularly around the location *Nedre Aalvik*, located south in the study area. At this location, the spatio-temporal and the two field model predict larger values of annual precipitation than the one field model. The difference from the one field model is around 600 mm/year on average over our study period.

As *Nedre Aalvik* is the location where the models were most different in terms of posterior mean, we take a closer look at the predictions here. In Figure 3a we visualize the posterior median at *Nedre Aalvik* with 95% posterior prediction intervals. We don't know the true, underlying precipitation at *Nedre Aalvik* in 1988 and after 1989. However, in the years where we have data, *Nedre Aalvik* is among the three or four wettest locations according to Figure 1b. The posterior medians in Figure 3a show that this tendency is captured by the two field model and the spatio-temporal model: They both predict *Nedre Aalvik* to be among the wetter locations, also in years without data. The tendency is not captured by the one field model as it

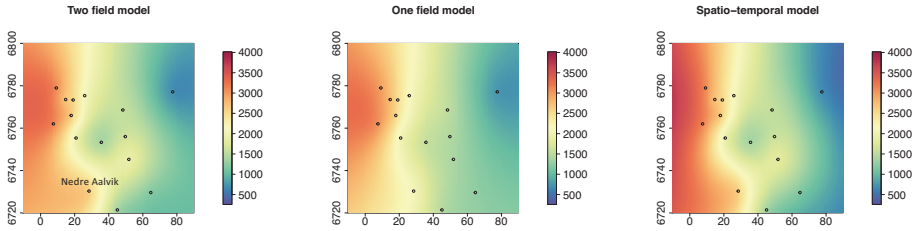
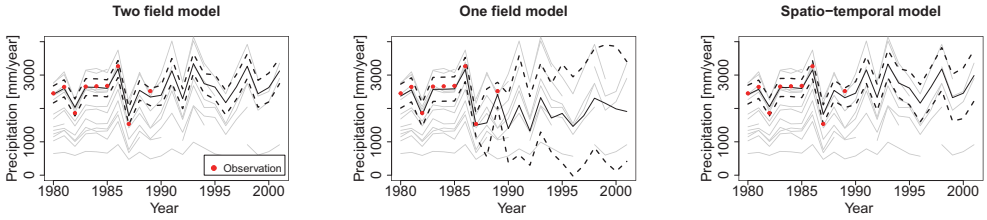


Figure 2: Average posterior mean precipitation [mm/year] around Voss over 22 years for the two field model (left), the one field model (middle) and the spatio-temporal model (right).

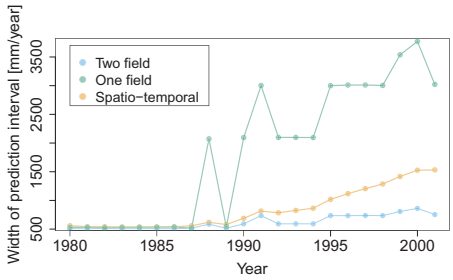
does not account for temporal dependencies. The results given by the more complicated models hence seem more reasonable.

In terms of predictive uncertainty, Figure 3a shows that the posterior uncertainties provided by the three models are similar for the years for which we have observations from *Nedre Aalvik*. In years without data, the posterior uncertainty of the two field model is lower than the posterior uncertainty of the one field model. This is reasonable when we look at the gray lines in Figure 3a that represent the underlying data: The Norwegian precipitation tend to follow strong, repeated spatial patterns. This can be seen by noticing that the time series are almost parallel, and the ranking between the stations (from wet to dry) is approximately unchanged over time. A relatively low uncertainty is hence realistic at locations where there are a few years of observations available.

Figure 3a further shows that also the spatio-temporal model is capable of giving a more realistic uncertainty than the one field model for our target location. However, the spatio-temporal model assumes that the posterior uncertainty increases over time. This is easier to see in Figure 3b where we visualize the width of the 95 % posterior prediction intervals for *Nedre Aalvik*. An uncertainty that increases as a function of time does not seem realistic for the annual precipitation in Norway as long as repeated weather



(a) Posterior median for *Nedre Aalvik* with 95 % prediction intervals. The gray lines mark the observations from the other precipitation gauges from Figure 1b.



(b) Width of the 95 % posterior prediction intervals at *Nedre Aalvik*.

Figure 3: The first plot shows the predicted precipitation at the location *Nedre Aalvik* for the two field model, the one field model and the spatio-temporal model. *Nedre Aalvik* has 9 observed values and 13 missing values between 1980 and 2001. The lower plot shows the corresponding widths of the 95 % posterior prediction intervals.

patterns are present. The observed data pattern suggests that the uncertainty model provided by the two field model best reflects the long-term behavior of annual precipitation in Norway.

Furthermore, the three models have different interpretations of future predictions. We illustrate this in Figure 4 where we have predicted runoff for 20 future years (2002-2021) for the location *Granvin* from Figure 1a. Mark that there is only one annual observation available from *Granvin* (1980).

First of all, the one field model does not support future predictions as it is a model only constructed for spatial interpolation. Next, considering the results from the spatio-temporal model, we again see that its posterior uncertainty increases with time. We also see that its posterior median slowly moves towards the global intercept β_c as the time since the last *Granvin* observation increases. This is different from the two field model, as we see in Figure 4: For the two field model, a future prediction is given by the estimated mean annual precipitation at the target location. The latter is given by the posterior median (or mean) of the climatic part of the model $\beta_c + c(\boldsymbol{s})$. The year specific part of the model $\beta_j + x_j(\boldsymbol{s})$ has zero mean and median for unobserved years. Hence, in the two field model each location can have its own long-term median or mean that is given by $\beta_c + c(\boldsymbol{s})$. This way it avoids regression towards the global intercept β_c over time. When studying variables that are driven by repeated weather patterns, this might be a more intuitive model for the future posterior median, than the posterior median provided by the spatio-temporal model. Also mark that although the posterior median for the two field model is given by $\beta_c + c(\boldsymbol{s})$ for a future year, all model components $\beta_c + c(\boldsymbol{s}) + \beta_j + x_j(\boldsymbol{s})$ contribute to the posterior uncertainty.

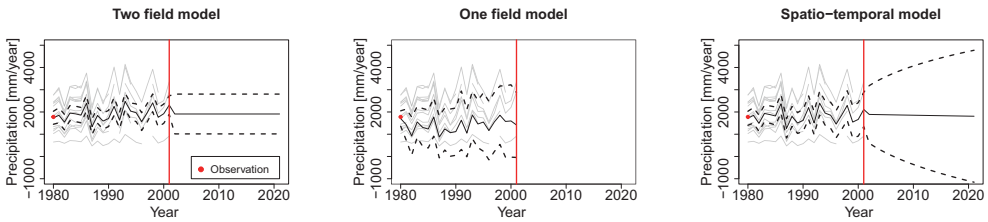


Figure 4: Posterior median with 95 % prediction intervals at the location *Granvin* that only has 1 annual observation of precipitation and 21 missing observations between 1980 and 2001. The gray lines mark the observations from the other precipitation gauges from Figure 1b. In this plot we have also included predictions for *future* years, i.e. for years after 2001 where we don't have data. The red vertical line marks the difference between the years where we have data and future years.

3.3 Parameter inference for the Voss example

We now study the parameter estimates of the three models for the Voss case example. In Table 1, the posterior quantiles of the parameters are shown. The posterior quantiles of the parameters of the climatic field $c(\boldsymbol{s})$ and the replicated field $x_j(\boldsymbol{s})$ are interesting here: We see that the posterior median of the marginal standard deviation σ_c is approximately 5 times larger (1083 mm/year) than the posterior median of σ_x (219 mm/year) for the two field model. This means that most of the spatial variability of annual precipitation around Voss can be explained by weather patterns that are repeated over time, i.e. the climate in the study area. This is not surprising considering the time series in Figure 1b where we see that the ranking between the locations (from wet to dry) is approximately constant over time. Table 1 also shows that the posterior median range of the most dominating spatial field $c(\boldsymbol{s})$ is smaller than the posterior median spatial range of the other spatial field $x_j(\boldsymbol{s})$ (96 km and 194 km respectively).

The above parameter values can be linked to the prediction results we saw in Section 3.2. Because the climatic field dominates in the study area ($\sigma_c \gg \sigma_x$), one or a few annual observations can have a large impact on the predictions at a target location. This was seen in Figure 3a where the climatic field $c(\boldsymbol{s})$ captured how the observations from *Nedre Aalvik* varied relative to the other study locations. This made the two field model able to predict *Nedre Aalvik* among the three or four wettest locations, also in years without data from *Nedre Aalvik*. Note that the strong, repeated spatial patterns of precipitation also are reflected in the parameter a for the spatio-temporal model in Table 1. This parameter is a measure of temporal dependency, and we see that it is close to 1. This means that the spatial pattern of annual precipitation is almost unchanged from one year to another.

Table 1: Posterior medians and (0.025 quantiles , 0.975 quantiles) for the parameters of the three models for the motivational example.

Parameter [unit]	Two field model	One field model	Spatio-temporal model
ρ_c [km]	96 (40,245)	-	-
σ_c [mm/year]	1081 (573,2144)	-	-
ρ_x [km]	195 (149,266)	89 (66,115)	164 (109, 244)
σ_x [mm/year]	219 (180,273)	993 (854,1159)	1559 (1002,2474)
β_0 [mm/year]	1722 (-1191,3794)	1818 (1433,2205)	1359 (-2673,4813)
τ_β [1/(mm/year)^2]	74 (10,1632) $\cdot 10^{-6}$	69 (6,3233) $\cdot 10^{-6}$	11 (4,31) $\cdot 10^{-6}$
a [1]	-	-	0.99 (0.98,1)
τ_ϵ [1/(mm/year)^2]	88 (65,118) $\cdot 10^{-6}$	172 (40,1652) $\cdot 10^{-6}$	75 (50,108) $\cdot 10^{-6}$

4 Simulation study: A more formal exploration of the two field model's properties

To explore the two field further, we present some simulation examples that were done to understand the relationship between the parameters of the two field model and its predictive performance. These are presented in Section 4.1 and Section 4.2. Next, in Section 4.3, we perform a simulation study to investigate if the two field model's parameters are identifiable.

4.1 Predictive performance and parameter values

In the illustrative example from Section 3, we connected the prediction results of the two field model to the strong climate around Voss ($\sigma_c \gg \sigma_x$). We now investigate the connection between the parameter values and the predictive performance of the model more formally, through a simulation study. This let us test other parameter settings, e.g. settings where $\sigma_c \ll \sigma_x$, and different observation designs. In our simulation study, we are particularly interested in investigating the model's ability to perform accurate predictions when there are a few years (or time units) of data available from the target location, such as for *Nedre Aalvik* and *Granvin* in the Voss example. It is in a setting like this the two field model should have its benefits,

according to existing results in Roksvåg et al. (2020b).

Experimental set-up

Table 2: Parameter used for simulating data from the two field model. For all 15 parameter configurations we use $\beta_c = 2$, $\tau_\beta = 5$ and $\tau_\epsilon = 9$, while the parameters below are tested for the spatial fields $x_j(\boldsymbol{s})$ and $c(\boldsymbol{s})$. The last column shows the percentage of the spatial variability that can be explained by the climatic spatial field.

Parameter set	ρ_c	ρ_x	σ_c	σ_x	$\sigma_c^2/(\sigma_c^2+\sigma_x^2)$
1	50	100	0.2	0.8	0.06
2	50	100	0.5	0.8	0.28
3	50	100	0.5	0.5	0.50
4	50	100	0.8	0.5	0.72
5	50	100	0.8	0.2	0.94
6	100	100	0.2	0.8	0.06
7	100	100	0.5	0.8	0.28
8	100	100	0.5	0.5	0.50
9	100	100	0.8	0.5	0.72
10	100	100	0.8	0.2	0.94
11	100	50	0.2	0.8	0.06
12	100	50	0.5	0.8	0.28
13	100	50	0.5	0.5	0.50
14	100	50	0.8	0.5	0.72
15	100	50	0.8	0.2	0.94

For this simulation study, we simulate data from the two field model from Section 2.1 and we use the same values for τ_β, τ_ϵ and β_c for all experiments. These values are set to $\beta_c = 2$, $\tau_\beta = 5$ and $\tau_\epsilon = 9$, and are chosen based on values that are realistic according to a study of annual runoff in the Voss area of Norway (Table 1 in Roksvåg et al. (2020a)), where runoff was given in m/year. For the parameters of the two GRFs ρ_c, ρ_x, σ_c and σ_x, different values are tested. The parameters are chosen such that the replicated field $x_j(\boldsymbol{s})$ dominates for some experiments ($\sigma_c \ll \sigma_x$), the climatic field $c(\boldsymbol{s})$ for others ($\sigma_c \gg \sigma_x$), and for some experiments the two spatial fields are equally dominant with the same marginal variance ($\sigma_x = \sigma_c$). Similarly, we

also do experiments with $\rho_c > \rho_x$, $\rho_c < \rho_x$ and $\rho_c = \rho_x$ for all configurations of marginal variances. Recall that ρ_c and ρ_x are the ranges of the climatic spatial field and the replicated spatial field respectively. Likewise are σ_c and σ_x the marginal standard deviations of the two fields. In total, we simulate from 15 configurations of marginal variances and ranges, and the parameters configurations are shown in Table 2.

For the assessment of predictive performance, we generate 50 simulated datasets for each of the 15 parameter sets in Table 2. Each of the 50 datasets has one climatic spatial field $c(\boldsymbol{s})$ and ten replicated fields $x_j(\boldsymbol{s})$ $j = 1, .., r$ with $r = 10$. We simulate data for the 15 locations from Figure 1a, i.e. the locations of the 15 precipitation gauges around Voss in Norway. We use this set-up such that the distribution of observation locations is realistic.

In the experiment, we make predictions of $\eta_j(\boldsymbol{s})$ for $j = 1, .., 10$ at three of the locations in Figure 1a: Voss, Brandset and Aurland. Voss and Brandset are located in the middle of the study area and have several close neighboring stations, while Aurand is located a bit further from the other observation locations. We predict $\eta_j(\boldsymbol{s})$ at either Voss, Aurland or Brandset by using observations from the remaining 14 locations from $j = 1, ..10$, and we explore two settings (S1 and S2):
S1: All the observations from Voss, Brandset or Aurland are removed from the dataset. This is equivalent to leave-one-out cross-validation.
S2: All the observations from Voss, Brandset or Aurland are removed from the dataset, except one. That is, there is one annual observation from the target location included in the observation likelihood. This is drawn randomly from the ten observations available.

We perform S1 and S2 for the one field model from Section 2.2 and the two field model from Section 2.1 to see how they differ in predictive performance for different parameter values. The spatio-temporal model from Section 2.3 is omitted from this larger simulation study due to its computational complexity. We perform S1 to explore if the two field model is well behaved for a standard spatial interpolation setting with totally unobserved target

locations. S2 is interesting because it is for locations that have a couple of observations available the two field model should have its benefits, begin able to transfer information across time through $c(\boldsymbol{s})$. The prior distributions we use for the simulation study, are similar to the ones used for the illustrative example from Section 3 (only rescaled to give predicted values in m/year instead of mm/year).

In the evaluation of the predictive performance of the one- and two field model, we use the root mean squared error (RMSE). We compute the RMSE for each of the 3 target locations (Voss, Brandset, Aurland) as

$$\text{RMSE}_i^{(b)} = \sqrt{\frac{1}{r}\sum_{j=1}^{n}(\eta_j^{(b)}(\boldsymbol{s}_i) - \hat{\eta}_j^{(b)}(\boldsymbol{s}_i))^2}, \tag{7}$$

where $\eta_i^{(b)}(\boldsymbol{s}_i)$ is the true simulated value and $\hat{\eta}_i^{(b)}(\boldsymbol{s}_i)$ is the corresponding predicted value for location $\boldsymbol{s}_i$, year j and simulation $b = 1, ...50$. The posterior mean is used as the predicted value. The RMSE is computed for each target location for each of the 50 simulated dataset, i.e. there are 150 RMSE evaluations in total per parameter configuration. As a summary score, the average $\text{RMSE}_i^{(b)}$ over these 150 experiments is used, which we denote $\overline{\text{RMSE}}$. Low RMSE values correspond to an accurate predictions.

Results

Figure 5 summarizes the results from the simulation study of the predictive performance: In this figure, the difference in $\overline{\text{RMSE}}$ between the two field model and the one field model is plotted for the different parameter sets. The results show that the two field model only gives a slightly lower $\overline{\text{RMSE}}$ than the simple model for S1, i.e for the leave-one-out cross-validation setting where no observations from the three target locations are included in the likelihood. This holds for all parameters configurations. A low difference in $\overline{\text{RMSE}}$ between the model types is as expected here. Having information

about the climatic spatial field $c(\boldsymbol{s})$ does not affect the predictive performance considerably when full data series from the observation locations are available. However, when there is one observation available from the target location, as in setting S2, the climatic spatial field does matter: In Figure 5 we see that as the fraction $\sigma_c^2/(\sigma_c^2+\sigma_x^2)$ increases, the difference in $\overline{\text{RMSE}}$ between the one field- and the two field model increases. Particularly when the climatic field's range is low (ρ_c=50) this is apparent. This is intuitive: If most of the spatial variability can be explained by the climate ($\sigma_c \gg \sigma_x$), and the climatic range ρ_c is low, it is less to learn from the neighboring stations compared to when the climatic range is large. One observation from the target location will thus have a large impact on $c(\boldsymbol{s})$ and consequently also a large (positive) impact on the predictive performance for the target location in the two field model. The one field model does not have this property as it does not account for temporal dependencies.

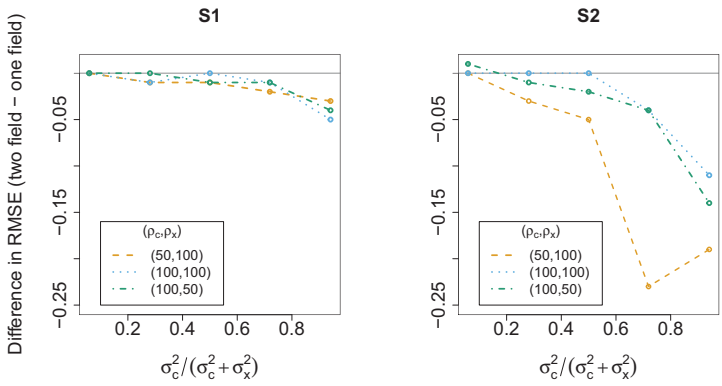


Figure 5: Difference in $\overline{\text{RMSE}}$ for predictions at Voss, Brandset and Aurland (two field model - one field model). More negative means that the two field model performs better. For S1, there are 0 observations from the target location in the likelihood, while for S2 there is 1 observation available out of 10. Predictions are performed for 50 datasets, each with one climatic field and $j = 1, ..10$ replicated fields.

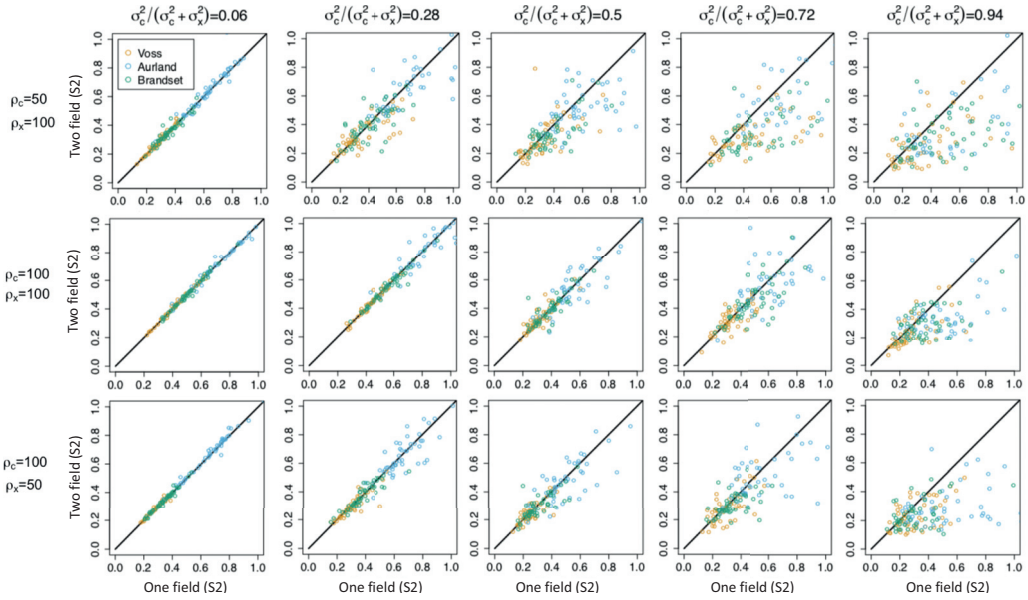


Figure 6: $\text{RMSE}_i^{(b)}$ for Voss, Brandset and Aurland for the one field model (x-axis) and the two field model (y-axis) for 50 simulated datasets for setting S2, i.e when one observation from the target location is included in the likelihood. Voss and Brandset are located in the middle of the study area in Figure 1a, while Aurland is located further away from the other observation locations. The two field model gives a better predictive performance than the one field model for points that are located under the black line.

In Figure 6 a more detailed comparison of the $\text{RMSE}_i^{(b)}$ between the two models is shown for the S2 case. We see that the models are approximately equally good when $\sigma_c < \sigma_x$. As we move rightwards in the plot, i.e as the fraction $\sigma_c^2/(\sigma_c^2 + \sigma_x^2)$ increases and the climatic effects become more dominating, the difference between the one field- and the two field model increases. The main trend is that the $\text{RMSE}_i^{(b)}$ is lower for the two field model than for the one field model, but we also see that the one field model "wins" sometimes. The larger variability in predictive performance right-

wards in the plot can be explained by the fact that if one of the models give a poor prediction for one year j, it typically also gives poor predictions for other years $j = 1, ..r$, as long as climatic effects are dominating. That is, if we fail to characterize the underlying climate, the same systematic error is performed over time. This happens both for the one field model and the two field model. However, the two field model tends to miss less often than the one field model on average, according to Figure 5 and Figure 6.

4.2 Simulation study from Roksvåg et al. (2020a)

The results from the above simulation study are related to the simulation study presented in Roksvåg et al. (2020a). Here, the connection between the two field model's parameters and systematic prediction bias was explored. It was shown that when the climatic field dominates over the replicated fields ($\sigma_c \gg \sigma_x$), there is a large probability of obtaining systematic biases for a prediction location over time. This holds as long as the observation set-up is the same. The reason is that when climatic effects dominate, the study area is driven by repeated weather patterns and the spatial pattern is approximately constant for $j = 1, ..r$. Hence, when taking a new set of measurement for a new time unit $r+1$, we don't really get any new information about the underlying spatial process unless we change our observation locations. Consequently, we also tend to do systematic prediction errors over time for unobserved locations, until new measurement locations are included in the likelihood. The prediction errors can be particularly severe if the study area has a low climatic range ρ_c and a large marginal variance σ_c. In this situation the information gain from neighboring locations is low. This is a similar conclusion as the one obtained from Figure 6 in the above simulation study.

4.3 Parameter identifiability

The two above simulation studies demonstrate that the parameter values of the two field model can give us information about the predictability of the target variable (Figure 5 and Figure 6). A next step is to verify that we actually can trust our parameter estimates. Motivated by this, we present a simulation study on the parameter identifiability of the two field model to make sure that it is well behaved. In particular, we are interested in investigating if we are able to separate the climatic spatial field from the replicated spatial field. In order to conclude that the two fields are separable, we require that we are able to correctly identify the parameters of the two spatial fields $c(\boldsymbol{s})$ and $x_j(\boldsymbol{s})$, i.e ρ_c, σ_c, ρ_x and σ_x.

Experimental set-up

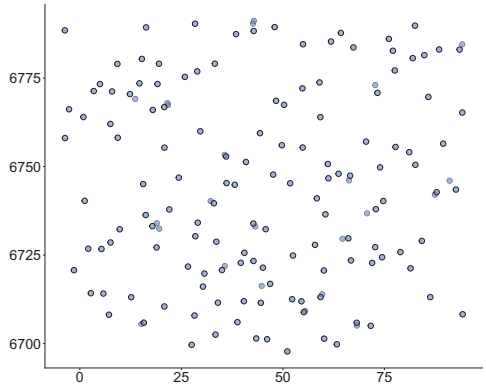


Figure 7: The 150 sampled study locations used for the simulation study that was carried out to investigate the identifiability of the model parameters.

In our exploration of identifiability of the parameters, we use a study area of size approximately 100×100 and we generate observations for 150 randomly drawn locations. The study area is shown in Figure 7. For all 15 parameter

combinations in Table 2, we generate 100 datasets from the two field model, for which each dataset has $r = 10$ replicates of the replicated field $x_j(\boldsymbol{s})$ $(j = 1, ..., 10)$ and one climate $c(\boldsymbol{s})$. That is, we use the same configuration of parameters as in the experiment presented in Section 4.1. We also use the same priors.

Identifiability results

Table 3 shows the average posterior medians for the parameters in the simulation study. The results show that the parameter τ_β is difficult to identify correctly. This is the precision of the year specific intercept β_j. For τ_β, the posterior median is often more than twice as large as the true underlying parameter. However, transforming the precision estimates to standard deviations $1/\sqrt{\tau_\beta}$, the predicted values are not that far away from the true standard deviation $1/\sqrt{5}$.

Table 3 further shows that the climatic range ρ_c is a bit hard to identify for some of the parameter configurations (Parameter set 1, 6 and 11). These are parameter configurations where the climatic spatial field has small spatial variability ($\sigma_c = 0.2$) compared to the replicated spatial field ($\sigma_x = 0.8$). It makes sense that the range ρ_c can be difficult to identify when the spatial field $c(\boldsymbol{s})$ only has a marginal impact on the final model: Some of its variability might be absorbed as random noise. It also makes sense that it is more difficult to identify ρ_c than ρ_x as we have several replicates of the replicated field $x_j(\boldsymbol{s})$, but only one realization of the climate $c(\boldsymbol{s})$. However, the estimated values for ρ_c for parameter sets 1 ,6 and 11 are not really that far away from the true range (around 10 units which is 10-20 % of the true range).

For the remaining parameters, the posterior medians are very close to the true underlying values. This indicates that the model is able to separate the two spatial fields $x_j(\boldsymbol{s})$ and $c(\boldsymbol{s})$ from each other. However, we have not investigated how our prior specification affects our results. For this simulation study we used the same priors for σ_c and ρ_c as for σ_x and ρ_x respectively

(see Section 3.1). We also used the same priors across experiments, regardless of the true simulated values of the parameters. With this in mind, it is a good sign that the spatial field parameters seems to be identifiable for all the tested parameter configurations. The parameter estimates don't look considerably drawn towards the underlying priors.

Table 3: Average posterior medians for the simulation study for parameter set 1-15. Each estimate is an average of 100 simulations. The number in parenthesis is the difference between the true value of the parameter and the estimated value (average posterior median minus true value). Thus, a positive value implies overestimation while a negative value implies underestimation. If the over- or underestimation is particularly high, the number in parenthesis is marked in red.

Param. set	ρ_c	ρ_x	σ_c	σ_x	τ_β	τ_ϵ	β_c
1	63.0 (+13)	101.7 (+1.7)	0.24 (+0.04)	0.80	10.5 (+5.5)	9.0	1.9 (-0.1)
2	52.8 (+2.8)	101.8 (+1.8)	0.52 (+0.2)	0.81 (+0.01)	12.6 (+7.6)	9.0	2.0
3	51.8 (+1.8)	99.3 (-0.7)	0.51 (+0.01)	0.49 (-0.01)	10.4 (+5.4)	9.1 (+0.1)	2.0
4	53.0 (+3.0)	100.9 (+0.9)	0.84 (+0.04)	0.50	12.6 (+7.4)	9.0	2.0
5	52.6 (+2.6)	107.3 (+7.3)	0.83 (+0.03)	0.20	7.0 (+2.0)	9.0	2.0
6	88.4 (-11.6)	100.3 (+0.3)	0.18 (-0.02)	0.80	10.7 (+5.7)	9.0	1.9 (-0.1)
7	99.6 (-0.4)	101.3 (+1.3)	0.50	0.81 (+0.01)	12.1 (+7.1)	9.1 (+0.1)	1.9 (-0.1)
8	97.9 (-2.1)	101.0 (+1.0)	0.49 (-0.01)	0.50	10.9 (+5.9)	9.1 (+0.1)	1.9 (-0.1)
9	96.6 (-3.4)	99.4 (-0.6)	0.77 (-0.03)	0.50	14.5 (+9.5)	9.0	1.9 (-0.1)
10	98.3 (-1.7)	100.0	0.78 (-0.02)	0.20	6.5 (+1.5)	8.9 (+0.1)	1.9 (-0.1)
11	88.7 (-11.3)	51.8 (+1.8)	0.21 (+0.01)	0.81 (+0.01)	15.6 (+10.6)	8.9 (-0.1)	1.9 (-0.1)
12	94.3 (-5.7)	51.7 (+1.7)	0.46 (-0.04)	0.81 (+0.01)	14.7 (+9.6)	9.0	2.0
13	95.6 (-4.4)	53.0 (+3)	0.48 (-0.02)	0.51 (+0.01)	8.0 (+3)	9.0	1.9 (-0.1)
14	96.7 (-3.3)	51.1 (+1.1)	0.75 (-0.05)	0.50	7.0 (+2)	9.0	2.0
15	108.7 (+8.7)	52.9 (+2.9)	0.83 (+0.03)	0.20	5.9 (+0.9)	9.0	1.9 (-0.1)

5 Recommendations: When to use a two field model?

We have presented one illustrative case example and a simulation study, which should give some intuition about the two field model. We now describe an ideal dataset for the two field model and propose some model applications.

5.1 Data structure

The Norwegian annual precipitation dataset in Figure 1 was chosen for the illustrative example because it was particularly suitable for the two field model. The reason is that the Norwegian annual precipitation follows repeated spatial patterns over time due to orographic precipitation that is driven by the country's topography. The repeated spatial patterns can be recognized from the time series in Figure 1b by that the time series from the different locations almost are parallel. The ranking between the stations, from wet to dry locations, is also approximately constant over time. This spatial trend can be transferred across time through the spatial field $c(\boldsymbol{s})$, which is the dominating model component for a spatial pattern like this.

Borrowing strength in time is particularly important if the target location has data available for some time units, but lacks data for most of the study period. If the climate is strong, only one (annual) observation could be enough to see large improvements in the predictability of a target variable, according to the results in Roksvåg et al. (2020b). This conclusion is valid in spite of large spatial variability. This was also indicated in the illustrative example, where only a few data points from *Nedre Aalvik* were needed in order to capture that this was among the wetter locations in the dataset (Figure 3a). Hence, several missing values is another property of the dataset in Figure 1 that makes the Norwegian annual precipitation data particularly suitable for using the two field model. For such datasets the two field model makes a difference compared to a one field model.

Finally, we note that there is no clear temporal dependency structure in the Voss data in Figure 1: Considering any location, it does not look like year j depends on year $j-1$. The two field model is most useful for variables that follow spatial patterns that are approximately constant over time, i.e. when there are no apparent increasing or decreasing temporal trend. If there is a decreasing or increasing temporal trend in the data, the two field model is able to give reasonable interpolation results, but not reasonable future predictions. An example related to this is shown later (see Example 2 in

Section 6).

Comparing the two field model to the alternative models, the one field model is not able to capture repeated spatial trends over time. The spatio-temporal model is on the other hand able to capture the spatial patterns, but is considerably slower than the two field model. It also provides regression towards the global intercept β_c (Figure 4) and assumes increasing uncertainty over time which is not necessarily realistic for a dataset like the one in Figure 1b. However, if there in fact are temporal dependencies in data (in addition to the underlying climate), a spatio-temporal model is the best option among the three discussed models, at least if future predictions are of interest.

We emphasize that if there are no repeated spatial patterns over time in the data, the two field model can still be used. However, in this case $\sigma_x \gg \sigma_c$ and the replicated field $x_j(\boldsymbol{s})$ dominates over the climatic field $c(\boldsymbol{s})$. The two field model will hence provide similar results as the one field model, as shown in Figure 5 and Figure 6, but at a higher computational cost. The simulation study also showed that the two field model and the one field model perform similarly when making predictions for locations that are totally unobserved. It is at locations where there are at least one or a few (annual) observations available that the two field model outperforms the one field model. Particularly will the two field model perform well compared to the one field model if the spatial variability σ_c is high, and the spatial range ρ_c is low, as shown in Figure 6. For such parameter configurations, the information gain from neighboring observations is low and a tiny piece of information from an otherwise unobserved location can be extremely valuable.

Based on the above discussion and the work in Roksvåg et al. (2020a) and Roksvåg et al. (2020b), we summarize how an ideal dataset for the two field model should look like.

Indications of a situation where using the two field model is suitable:
(i) The target variable follows a (constant) spatial pattern that is (partially) repeated over time. Mark that the year-to-year variability still can be large: It is the *spatial pattern* that should be stable.
(ii) For any given observation location in the dataset, there is no apparent temporal dependency. This is most relevant if the user is interested in future predictions and not only spatial interpolation.
(iii) If the user is primarily interested in predictions (not inference), the dataset should have some missing values. An ideal target location has a few observations available, but lack observations for most of the study period. However, long time series from nearby locations should be available.

The dataset from Figure 1 has all the above properties and can be used as a reference dataset.

5.2 Suggested applications

Based on Roksvåg et al. (2020a,b) and the above examples, we also suggest some areas of use for the two field model.

Infill of missing data: Infill of missing data in environmental datasets is the two field model's primary area of use, i.e. to fill in missing observations for locations where there only are short data records available, but where there are longer time series of data available from neighboring locations. Here, the model has major advantages compared to other statistical models, as demonstrated in the simulation study. The user should note that the data included in an analysis don't need to be consecutive: If there for example are some data available from 1940, this year can be modeled together with data from 1980-1990 to include more observation locations. The two field model was successfully used for infill of missing values for a dataset of Norwegian annual runoff in Roksvåg et al. (2020b).

Modeling long-term averages: In many applications it is of interest to study the long-term average of an environmental variable, typically over a 30 year period, as this defines a climate (WMO, 1992). For the two field model the predicted long-term average of the variable can be read directly from the distribution of $\beta_c + c(\boldsymbol{s})$. The long-term average is also what the two field model gives as the (posterior mean) future prediction (see Figure 4). Based on this, the two field model should be particularly suitable for modeling long-term averages, such as mean annual/monthly precipitation, mean annual/monthly runoff and mean annual/monthly temperature. Long-term averages also tend to respond slowly to climatic changes, and it is often a reasonable assumption that the spatial pattern of an environmental variable of interest is repeated annually over e.g. a 10 or 20 year period.

Understanding biases in environmental modeling: The two field model quantifies how much of the spatial variability that can be explained by climatic trends through the parameters of the spatial fields σ_x, σ_c, ρ_x and ρ_c. This can be a useful tool for understanding prediction biases in environmental modeling, as described in Section 4.2 and further documented Roksvåg et al. (2020a). More specifically, if climatic effects dominate and the range ρ_c is low, systematic prediction biases can be expected over time for many of the prediction locations.

The value of information (VoI): The increase in predictive performance associated with one (annual) observation from the target location can be large when $\sigma_c \gg \sigma_x$, but low if $\sigma_c \ll \sigma_x$. Hence, the model and its parameters can be used to motivate the installation or shutdown of a measuring station. If $\sigma_c \gg \sigma_x$ the installation of a new measuring station could improve the predictions at this location short time after installation. The above example is related to decision theory and assessing the value of information connected to gathering new observations (Eidsvik et al., 2015). Using the two field model for finding an optimal observation design could be an interesting topic for further research.

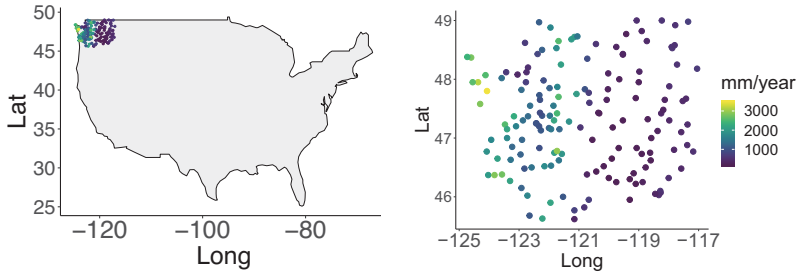
6 Other case examples

We have demonstrated the two field model for simulated data and for Norwegian annual precipitation. Existing work also shows that the two field model is useful for modeling annual runoff in Norway (Roksvåg et al., 2020a,b). In this section we use the two field model to model four other target variables with corresponding study areas. These target variables follow similar spatial trends as the Norwegian annual precipitation data in Figure 1b. In the first example we also identify a potential model weakness.

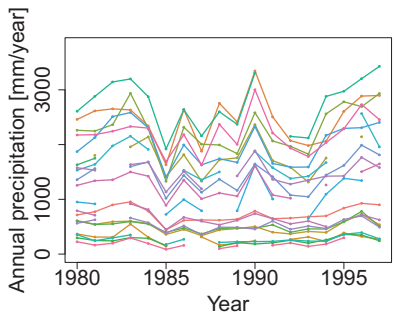
Example 1: Annual precipitation in the state of Washington

Figure 8 shows annual precipitation data from 1980-1997 from the state of Washington in the US. The data were downloaded from http://www.image.ucar.edu/Data/US.monthly.met/. This dataset was also used in Fuglstad et al. (2015) to fit a non-stationary precipitation model. The time series in Figure 8b indicates that there are climatic trends in the data that can be exploited. Furthermore, there are several missing values in the Washington dataset, which makes the data suitable for a further analysis with the two field model.

A potential problem with the Washington dataset however, is that there are quite large differences in temporal variance across the study area. This can be seen in Figure 8b: Precipitation stations with low measurements of annual precipitation tend to have lower temporal variance than precipitation stations with high measurements. This is probably a common property for larger, environmental datasets, and it is not clear how the two field model handles this. To investigate this further, we did an experiment where we predicted precipitation for three of the stations in Washington. We chose three locations that were located in areas with different temporal variability (low, medium and high). The stations also only had 1 or 2 annual precipitation observations available (out of 18 years), i.e. there were many missing values that we could predict.



(a) Average precipitation from 140 locations in Washington state from 1980-1997.



(b) Time series from 21 of the locations.

Figure 8: Annual precipitation from Washington state from 1980-1997. The distances on the left map are given in degrees [°].

The two field model from Section 2.1 was fitted to the dataset, and gave the following posterior means for the parameters of the spatial fields: $\rho_c = 0.85°$ (long/lat), $\rho_x = 2.1°$ (long/lat), $\sigma_c = 865$ mm/year and $\sigma_x = 171$ mm/year. Hence, as $\sigma_c \gg \sigma_x$ most of the spatial variability can be explained by repeated weather patterns. This is similar to the Norwegian precipitation data. Further, the predictions for the three selected precipitation stations in Washington are visualized in Figure 9. The results indicate that the two field model is able to handle differences in temporal variability across the study area for years where we have data (1980-1997): There are clearly dif-

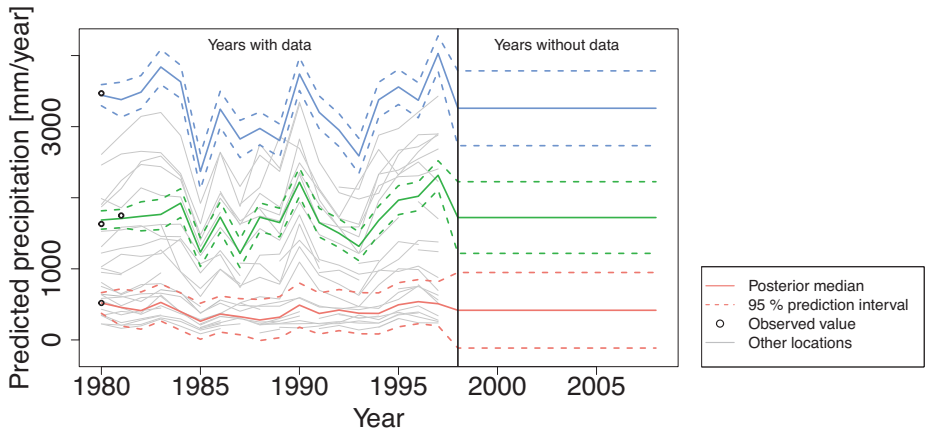


Figure 9: Posterior median annual precipitation with corresponding 95 % posterior prediction intervals for three precipitation stations in Washington. The black points are the observed values from the three target locations, while the gray lines are observations from some selected nearby locations. The two field model seem to be able to capture different temporal variances in years where we have data (1980-1997). However, the uncertainty model for future, unobserved years (1998-2008) is similar for the three target locations. This might not be realistic.

ferences in the temporal variability for the three precipitation stations, also for the predicted time series. Next, considering predictions for future years without data (1998-2008), the model seems to give reasonable predictions in terms of the posterior median (Figure 9). Recall that the posterior median for a future prediction is given by the climatic component $\beta_c + c(\boldsymbol{s})$ and can be interpreted as the long-term annual precipitation for the target location. However, considering the future prediction intervals, the width of these are almost equal for the three locations, i.e. the locations have similar posterior variances for 1997-2008. This might not be realistic and a potential model weakness for a dataset like this. To improve the uncertainty estimates for future predictions, a climatic field with a non-stationary spatial variance could be implemented. However, this represents a large increase in

the computational complexity of the problem. In addition, non-stationary fields are often difficult to identify without replicates (Ingebrigtsen et al., 2014, 2015), and we necessarily only have one realization of our underlying climate.

Example 2: Increasing trend

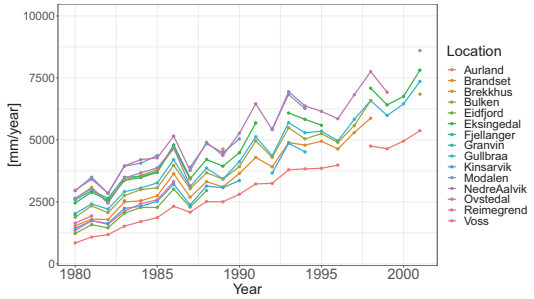


Figure 10: The data from Figure 1b with an increasing temporal trend added to them. We explore if the two field model is able to capture a pattern like this.

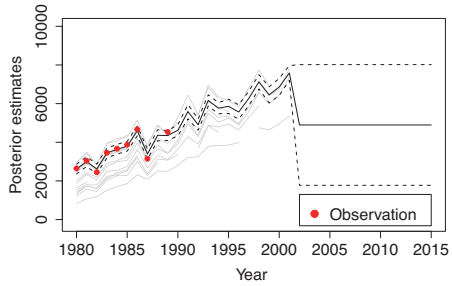


Figure 11: Posterior mean for *Nedre Aalvik* for 1980-2001 with 95 % posterior prediction intervals for the two field model. The gray lines marks the precipitation observations from the 14 remaining locations. The red line marks future years (after 2001).

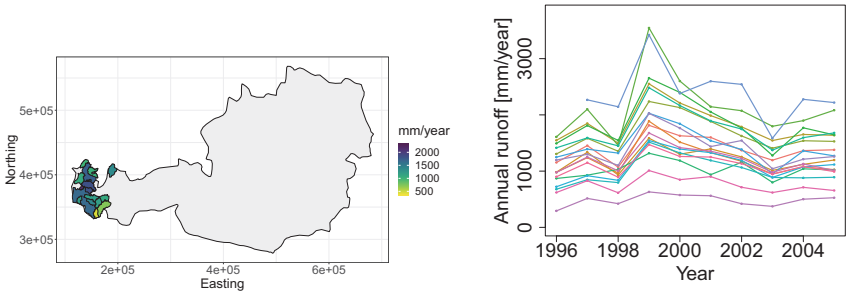
In this example we use the data in Figure 1 again, i.e. the precipitation data from Voss. However, we have added an increasing temporal trend to the data to see how the two field model handles this. The trend is simulated each year from a normal distribution with mean 200 mm/year and standard deviation 10 mm/year, and aggregated over the years $j = 1, ..r$. The data can be seen in Figure 10.

The two field model model was fitted to the data in Figure 10, and used to predict the target variable at the location *Nedre Aalvik* from Figure 1 for 1980-2001. The posterior means are shown in Figure 11. We see that the suggested framework is able to capture the increasing trend, also for years for which there are no measurements from *Nedre Aalvik*. This is possible due to the year specific intercept β_j. At the same time, the model is able to capture the underlying spatial pattern, i.e. the ranking between the observation locations: *Nedre Aalvik* is correctly predicted among the wetter locations in the study area. The posterior uncertainty also looks reasonable.

However, if future predictions are of interest (and not only spatial interpolation), a spatio-temporal model with an increasing trend is a better model. The two field model is not able to give reasonable predictions for future years when an increasing trend is present, as Figure 11 shows.

Example 3: Annual runoff in Austria

In Figure 12 we see time series of annual runoff from 1996-2005 for 18 catchments in the western part of Austria. The data are freely available at `ehyd.gv.at`. The time series in Figure 12b indicate that the runoff in western Austria is driven by spatial patterns that are repeated annually. This is confirmed by fitting the two field model to the data (with the catchment centroids as observation locations): The posterior mean of σ_x is 116 mm/year while the posterior mean of σ_c is 604 mm/year. As for the Norwegian annual precipitation, the posterior estimate of ρ_c is smaller than the posterior estimate of ρ_x with posterior means 13 km and 51 km respectively.

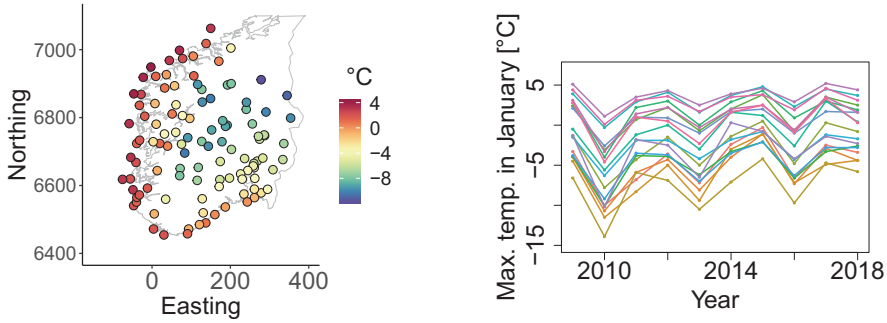


(a) 18 catchments in Austria and their average annual runoff from 1996-2005.

(b) Time series for the 18 Austrian catchments.

Figure 12: Annual runoff data from Austria from 1996-2005. The distances on the left map are given in meters.

Example 4: Maximum monthly temperature in Norway



(a) The average maximum temperature in January for 2009-2018.

(b) Time series from 18 of the locations.

Figure 13: Maximum temperature for January for southern Norway (2009-2018) for 111 locations. The distances on the left map are given in km.

Figure 13 shows the average maximum temperature in January for southern

Norway for 2008-2019. The data were downloaded at `eklima.no`, a web page maintained by the Norwegian Meteorological institute. The time series in Figure 1b show that there also for this variable are repeated spatial trends as the time series vary systematically compared to each other. Fitting the two field model to this data gives posterior means $\rho_x = 181$ km, $\sigma_x = 0.9$ °C, $\rho_c = 111$ km and $\sigma_c = 4.5$ °C, i.e. most of the spatial variability can be explained by repeated weather patterns ($\sigma_c \gg \sigma_x$). Hence, this is another dataset that should be suitable for further analysis with the two field model.

7 Implementing the two field model in r-inla

Finally, we show how the two field model can be implemented in the `R` package `R-INLA` (`www.r-inla.org`). The main contribution of this section is to demonstrate how to specify the two spatial fields $x_j(\boldsymbol{s})$ and $c(\boldsymbol{s})$ and highlight the differences in the specification of these two. For a more in-depth presentation of the different `R-INLA` functions used in our implementation, we refer to Blangiardo and Cameletti (2015); Krainski et al. (2018). Example code and example data for the two field model are also available on `https://github.com/tjroksva/RunoffInterpolation`. Here, we use the annual runoff data from Roksvåg et al. (2020b) as example data and specify priors for all model parameters. Prior specification for all parameters are not covered in the code below.

Assume we are going to fit the two field model to the dataset in Figure 1c. We upload the precipitation data as follows (with unit m/year):

```
> library(INLA)
>
> precipdata=as.matrix(read.table("precipdata.txt",header=TRUE))/1000
> r=dim(precipdata)[1]
> n=dim(precipdata)[2]
> coords=as.matrix(read.table("precip_coords.txt",header=TRUE))
> y=as.vector(t(as.matrix(precipdata)))
```

where precipdata is a matrix organized as in Figure 1c with each column representing a location 1,...,n and each row representing a year or time unit 1,...,r. Missing values are specified as NA's. The matrix coords is a n x 2 matrix that contains the x- and y- coordinates of the measurement locations. These coordinates are given in the same order as the columns in precipdata. Furthermore, in r-inla the observations have to be stored in a vector, and we therefore make vector y. In this vector, all the observations from year 1 come first, then all the observations from year 2, all up to year r.

Next, we construct a suitable triangulation mesh for the spatial domain such that we can use the SPDE approach to spatial modeling:

```
> mesh=inla.mesh.2d(loc.domain=coords,cutoff=3,max.edge=c(5,15),
>      offset=c(20,30))
```

Here it can be useful to type mesh$n to see the number of mesh nodes in the triangulation. This number should be large enough to capture the spatial variability in the study area, but if it is too large, inference is slow. We refer to Blangiardo and Cameletti (2015) for more on mesh construction. To be able to transfer approximations from the mesh nodes to the target locations in coords, we make a projection matrix A. This is done by the command

```
> A=inla.spde.make.A(mesh,loc=coords,
>                      index=rep(1:n,times=r),repl=rep(1:r,each=n))
```

So far we have only used standard R-INLA commands. Now we show how to construct a replicated field $x_j(\boldsymbol{s})$ for year j=1,..r:

```
> spde.x=inla.spde2.pcmatern(mesh=mesh,alpha = 2,
>                            prior.range = c(10,0.1),
>                            prior.sigma = c(2,0.1))
> x.index=inla.spde.make.index(name="x.field",
>                              n.spde=spde.x$n.spde,n.repl=r)
```

Note that the PC prior for the spatial parameters is specified in prior.range and prior.sigma. The climatic field $c(\boldsymbol{s})$ is constructed sim-

ilarly:

```
> spde.c=inla.spde2.pcmatern(mesh=mesh,alpha = 2,
>                            prior.range = c(10,0.1),
>                            prior.sigma = c(2,0.1))
> c.index=inla.spde.make.index(name="c.field",
>                            n.spde=spde.c$n.spde,n.repl=r)
> c.index$c.field.repl=rep(1,times=mesh$n*r)
```

Both the replicated field represented by `x.index` and the climatic field represented by `c.index` are lists of three vectors. The first vector is similar for the two fields: `x.index$x.field` and `c.index$c.field` are given by `rep(1:mesh$n,times=r)`, i.e. they are vectors counting from `1` to the number of mesh nodes, `r` times. The second vector is also the same for the two fields: `x.index$x.field.group` and `c.index$c.field.group` are given by `rep(1,times=r*mesh$n)`, i.e. it is only a vector of ones.

It is the third vector that distinguishes the replicated field $x_j(\boldsymbol{s})$ from the climatic field $c(\boldsymbol{s})$. For the replicated field, the third vector `x.index$x.field.repl` is given by `rep(1:r,each=mesh$n)`. That is, the first `mesh$n` elements are `1`'s and the next `mesh$n` elements are `2`'s. This continues up to `r`. This way we mark the `r` replicates of the spatial field $x_j(\boldsymbol{u})$ from j=`1,..r`. For the climatic field however, the third vector is just a vector of `1`'s, i.e. `c.index$c.field.repl= rep(1,times=r*mesh$n)`. This is done to mark that we don't have any replicates of the climate (the climate is assumed constant), and that all mesh nodes represent the same realization (realization `1` or replicate `1`).

After making indices for the two spatial fields, we specify indices for the replicated intercept β_j as follows:

```
> iid.index=list(iid=rep(1,times=mesh$n*r),
>           iid.repl=rep(1:mesh$n,each=r))
```

Here, the vector `iid.repl` is specified similarly as the replicated field's index `x.index$x.field.repl`. This is to mark that we have `r` replicates of β_j.

At this point, we have specified the necessary indices for all model components. The indices and the data `y` are next inserted into a `stack` together with the projection matrix `A`:

```
> stack=inla.stack(data=list(y=y),
>                  effects=list(c(x.index,c.index,iid.index,intercept=1)),
>                  A=list(A),tag="obs")
```

The linear predictor $\eta_j(\boldsymbol{s}) = \beta_c + x_j(\boldsymbol{s}) + c(\boldsymbol{s}) + \beta_j$ can now be specified as follows

```
> formula = y~ -1 +intercept+
>          f(x.field, model=spde.x,replicate=x.field.repl)+
>          f(c.field,model=spde.c)+
>          f(iid,model="iid",replicate=iid.repl,fixed=FALSE)
```

and we run the analysis by writing

```
> res=inla(formula,family=c("gaussian"),
>       control.compute=list(config = TRUE),
>       data=inla.stack.data(stack),
>       control.predictor=list(A=inla.stack.A(stack),compute=TRUE))
```

To get the posterior mean and the posterior standard deviation on the same format as the data in Figure 1c, write:

```
> index.results=inla.stack.index(stack,tag="obs")$data
> postmean=matrix(res$summary.fitted.values[index.results,"mean"],
          byrow=TRUE,ncol=n,nrow=r)
> postsd=matrix(res$summary.fitted.values[index.results,"sd"],
        byrow=TRUE,ncol=n,nrow=r)
```

The parameter estimates are found in:

```
> res$summary.hyperpar
> res$summary.fixed
```

This concludes the implementation of the two field model. For reference we also include how the simple spatial model from Section 2.2 and the spatio-temporal model from Section 2.3 can be implemented by making some modifications in the code above.

One field model: To implement the simple spatial reference model, remove the component `f(c.field,model=spde.c)` from the `formula`. Also omit the specification of `spde.c` and remove the component `c.index` from the `stack`.

Spatio-temporal model: Do the same modifications as for the one field model. In addition, replace the `x.index` specification by the following:

```
x.index=inla.spde.make.index(name="x.field", n.spde=spde.x$n.spde,
    n.group=r)
```

Also replace the `f(x.field,model=spde.x,replicate=x.field.repl)` in the `formula` by:

```
f(x.field, model=spde.x,group=x.field.group,
  control.group=list(model="ar1")).
```

8 Conclusions

In this note, we have explored the properties of a geostatistical two field model for modeling spatio-temporal data. It is suitable for modeling environmental variables that follow repeated spatial patterns over time and for modeling long-term averages. The two field model was first presented in Roksvåg et al. (2020a,b) where it was used for interpolation of annual runoff. In this note we have answered some questions that were considered outside the scope of these articles: We have for example illustrated when the two field model is useful compared to other statistical models for spatial and spatio-temporal data, and shown that the most important model parameters are identifiable. Furthermore, we have illustrated the model for new target variables and given a tutorial on how the model can be implemented in `r-inla`. This represents new contributions compared to the work in Roksvåg et al. (2020a,b).

References

S. Banerjee, A.E. Gelfand, and B.P Carlin. *Hierarchical Modeling and Analysis for Spatial Data*, volume 101 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, 2004.

M. Blangiardo and M. Cameletti. *Spatial and Spatio-temporal Bayesian Models with R-INLA*. Wiley, 1st edition, 2015.

N. Cressie. *Statistics for spatial data*. J. Wiley & Sons, 1993.

J. Eidsvik, T. Mukerji, and D. Bhattacharjya. *Value of Information in the Earth Sciences: Integrating Spatial Modeling and Decision Analysis*. Cambridge University Press, 2015. doi: 10.1017/CBO9781139628785.

G-A. Fuglstad, D. Simpson, F. Lindgren, and H. Rue. Does non-stationary spatial data always require non-stationary random fields? *Spatial Statistics*, 14:505 – 531, 2015. ISSN 2211-6753. doi: https://doi.org/10.1016/j.spasta.2015.10.001.

G-A. Fuglstad, D. Simpson, F. Lindgren, and H. Rue. Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association*, 114(525):445–452, 2019. doi: 10.1080/01621459.2017.1415907.

A. Gelfand, P. Diggle, M. Fuentes, and P. Guttorp. Handbook of spatial statistics. *Chapman & Hall*, 01 2010. doi: 10.1201/9781420072884.

R. Ingebrigtsen, F. Lindgren, and I. Steinsland. Spatial models with explanatory variables in the dependence structure. *Spatial Statistics*, 8:20 – 38, 2014. doi: 10.1016/j.spasta.2013.06.002.

R. Ingebrigtsen, F. Lindgren, I. Steinsland, and S. Martino. Estimation of a non-stationary model for annual precipitation in southern Norway using replicates of the spatial field. *Spatial Statistics*, 14:338 – 364, 2015. doi: 10.1016/j.spasta.2015.07.003.

E. Krainski, V. Gómez-Rubio, H. Bakka, A. Lenzi, D. Castro-Camilo, D. Simpson, F. Lindgren, and H. Rue. *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA*. 09 2018. doi: 10.1201/9780429031892.

M.N.M. Lieshout. *Theory of Spatial Statistics: A Concise Introduction*. 03 2019. ISBN 9780429052866. doi: 10.1201/9780429052866.

F. Lindgren, H. Rue, and J. Lindström. 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)*, 73:423–498, 2011. doi: 10.1111/j.1467-9868.2011.00777.x.

T. Roksvåg, I. Steinsland, and K. Engeland. A geostatistical two field model that combines point observations and nested areal observations, and quantifies long-term spatial variability – a case study of annual runoff predictions in the voss area. *arXiv:1904.02519v4*, 2020a.

T. Roksvåg, I. Steinsland, and K. Engeland. Estimation of annual runoff by exploiting long-term spatial patterns and short records within a geostatistical framework. *Accepted for publication in Hydrology and Earth System Sciences*, 2020b.

H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 2005.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71:319–392, 2009. doi: 10.1111/j.1467-9868.2008.00700.x.

D. Simpson, H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, 32:1–28, 2017. doi: 10.1214/16-STS576.

WMO. *International meteorological vocabulary*. 1992.