A geostatistical framework for estimating flow indices by exploiting short records and long-term spatial averages - Application to annual and monthly runoff

Thea Roksvåg[1], Ingelin Steinsland[1], and Kolbjørn Engeland[2]

[1]Norwegian University of Science and Technology, NTNU, Department of Mathematical Sciences.
[2]The Norwegian Water Resources and Energy Directorate, NVE

Correspondence: Thea Roksvåg (thea.roksvag@ntnu.no)

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 missing values and short records of data. A key feature of the proposed framework is that several years of runoff are modeled simultaneously with two Gaussian random fields (GRFs): One that is common for all years under study and represents the runoff generation due to long-term climatic conditions, and one that is year specific. The climatic GRF learns how short records of runoff from partially gauged catchments vary relatively to longer time series from other catchments, and transfers this information across years. Another property, is that the model takes the nested structure of catchments into account such that the water balance is preserved for any point in the landscape. The framework is demonstrated by interpolation of annual and monthly runoff from around 200 catchments in Norway, and we compare it to Top-Kriging (interpolation method) and simple linear regression (method for exploiting short records). The results show that if the correlation between neighboring catchments is high, a model that considers several years of runoff simultaneously is considerably better at capturing large spatial variability than a model that treats each year of data separately.

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 assessment of reliability of water supply, while average 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. Average annual flow is also often used as an predictor for both low flow and high flow indices (Sælthun et al., 1997).

At locations with measurements, the streamflow indices can be estimated based on the 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 such flow indices, but a challenge is that 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

5 for runoff interpolation particularly suitable for datasets including data series of this type, more specifically runoff datasets including a mix of fully gauged and partially gauged catchments. 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

10 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 depends on the pair-wise distances between the n 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 nested catchment areas, and not

15 to point locations in space. Traditionally, this problem 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

20 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 sub-catchment is weighted more than information from a nearby non-overlapping catchment when performing runoff predictions in an ungauged catchment.

The common approach for exploiting short records of runoff data is to find one or several donor catchments with longer time series of runoff. The donor catchments are typically found based on runoff correlation, catchment similarity, or proximity

25 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), 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 mean annual runoff, using the mean annual runoff

30 is a natural choice (McMahon et al., 2013) whereas for low flow indices, low flow observations are used. The predictive performance of these methods is highly dependent on the correlation between the runoff observations in the target catchment and the donor catchment. If the correlation is high and the spatial variability of runoff is high, short records of data from the target catchment can be extremely valuable. In the remainder of the paper, we will denote high correlation between runoff observations in an area as *hydrological stability*.

In this article we propose two geostatistical Bayesian models that are constructed to fully exploit the statistical pattern stored in sparse dataset, i.e hydrological datasets with several missing elements and short records of data. Such datasets are common in hydrology. In our presentation of our two models, we develop a methodology for *annual* runoff, but we also argue that it is a flexible framework that can be adapted to other flow indices of interest, e.g monthly runoff. A key feature of the two suggested

5   models is that they simultaneously model 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 long-term spatial average runoff, or the climate in the study area. 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

10   capture this tendency. This way, the climatic spatial field learns how short records of runoff vary relatively to longer data series from other catchments. If there is no strong climatic trend present in the data, the annual GRF will dominate over the climatic GRF and short records from the target catchment(s) will have less impact on the results. Thus, our method represents a way for detecting hydrological stability and uses this to exploit short records and to perform runoff interpolation.

The first model we propose is denoted the areal model and is particularly suitable for mass-conserved hydrological variables.

15   It ensures that the water-balance is close to preserved for any point in the landscape (with some uncertainty), 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 runoff is similar to the one of Top-Kriging. Another benefit of this model is an accurate and realistic representation of the correlation structure between catchments that are nested. A similar model has already shown promising results for predictions of annual runoff around Voss in Norway (Roksvåg et al., 2019), but

20   it is not tested for a larger dataset or for other time-scales. This is done in this article.

As an alternative to the areal model we also propose a model that defines runoff as a point referenced process for which distances are measured between catchment centroids. This model does not consider preservation of the water-balance, but on the other hand it can be used for any environmental variable, and it is computationally faster than the areal model. This is more similar to what has been done traditionally in hydrology, and we denote this the centroid model. Both the areal model and the

25   centroid model have the ability to exploit hydrological 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 article is to present and evaluate this new geostatistical framework. In particular our goals are to: 1) Assess the predictive performances of our two new spatial models and compare the results to the predictive performance of Top-Kriging and/or simple linear regression for predictions in ungauged and/or partially gauged catchments.

30   2) Assess the added value of including short records.

3) Demonstrate the approach for a real practical problem by producing a runoff map for mean annual runoff for southern Norway based on all available data from a 30 year period (1981-2010).

By 1) to 3) we also illustrate the benefits of modeling several years of runoff data simultaneously with a joint climatic GRF compared to treating each year of data separately with only annual GRFs.

The rest of the article is structured as follows: In Section 2 we present the dataset we use that consists of annual and monthly runoff data from catchments in Norway. Then, we briefly introduce relevant statistical background theory and notation in Section 3.1. Next, the suggested model for annual runoff is presented in Section 3.2.1 - 3.2.6, and extended to monthly runoff in Section 3.2.7. Evaluation scores and experimental set-up are presented in Section 4, before the results are presented and

5   discussed in Section 5. Finally, we summarize the major findings in Section 6 and conclude in Section 7.

2   Study area

The study is carried out by utilizing a dataset from Norway provided by the Norwegian Water Resources and Energy Directorate (NVE). It consists of daily runoff data from around 450 gauged catchments, many of them nested, from 1970-2018.

In this paper we use our framework to predict annual and monthly runoff. To make a test dataset for cross-validation, the daily

10   runoff data were aggregated to monthly and annual runoff for the hydrological years 1996-2005. In Norway the hydrological year starts September 1th and ends August 31st. Only catchments with observations every day for the hydrological years 1996-2005 were included in the test data.The years 1996-2005 were chosen because full time series for a large number of catchments were available for these specific years. This leaves 195 catchments for testing with areas ranging from 7.5 km^2 to 18934 km^2. The median elevation of the catchments ranges from 85 to 1562 m a.s.l. Figure 1a and Figure 1c summarize the

15   annual data. We see a large spatial variability of runoff. The annual runoff (for individual years) ranges from 170 mm/year to 5500 mm/year, whereas the mean annual runoff ranges from 370 mm/year to 4300 mm/year, with the highest values of runoff in western Norway and more moderate values in east and north.

In western Norway some of the largest values of annual precipitation in Europe are measured. The large values of annual precipitation in western Norway are mainly caused by the orographic enhancement of frontal precipitation formed around

20   extratropical cyclones. The orographic enhancement is explained by the 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 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 processes explained above, creates the east-west pattern in Figure 1a that is prominent each year in the dataset. The topography results in a spatial pattern of runoff that is stable over years, and suitable for exploiting short

25   records of runoff. An example of the spatial stability for annual runoff in Norway is demonstrated in Figure 3b, which shows time series of annual runoff from seven catchments in south-western Norway. We see variation from year to year. Nevertheless, the seven lines are almost parallel, showing that the annual runoff is highly correlated between these catchments. This tendency is typical for annual runoff for many of the catchments in Norway.

Observed monthly runoff from a winter month dominated by snow accumulation (January), a spring month with snow

30   melting (April) and a summer month dominated by rain (June) was chosen for an analysis of monthly runoff. On a monthly scale, full data series from some extra catchments existed in addition to the 195 catchments in Figure 1a. Data from 216 catchments were available for January, while data from 213 catchments were available for April and June. The monthly runoff data from Norway are presented in Figure 2, and we see that the monthly runoff is the lowest in January and the highest in June
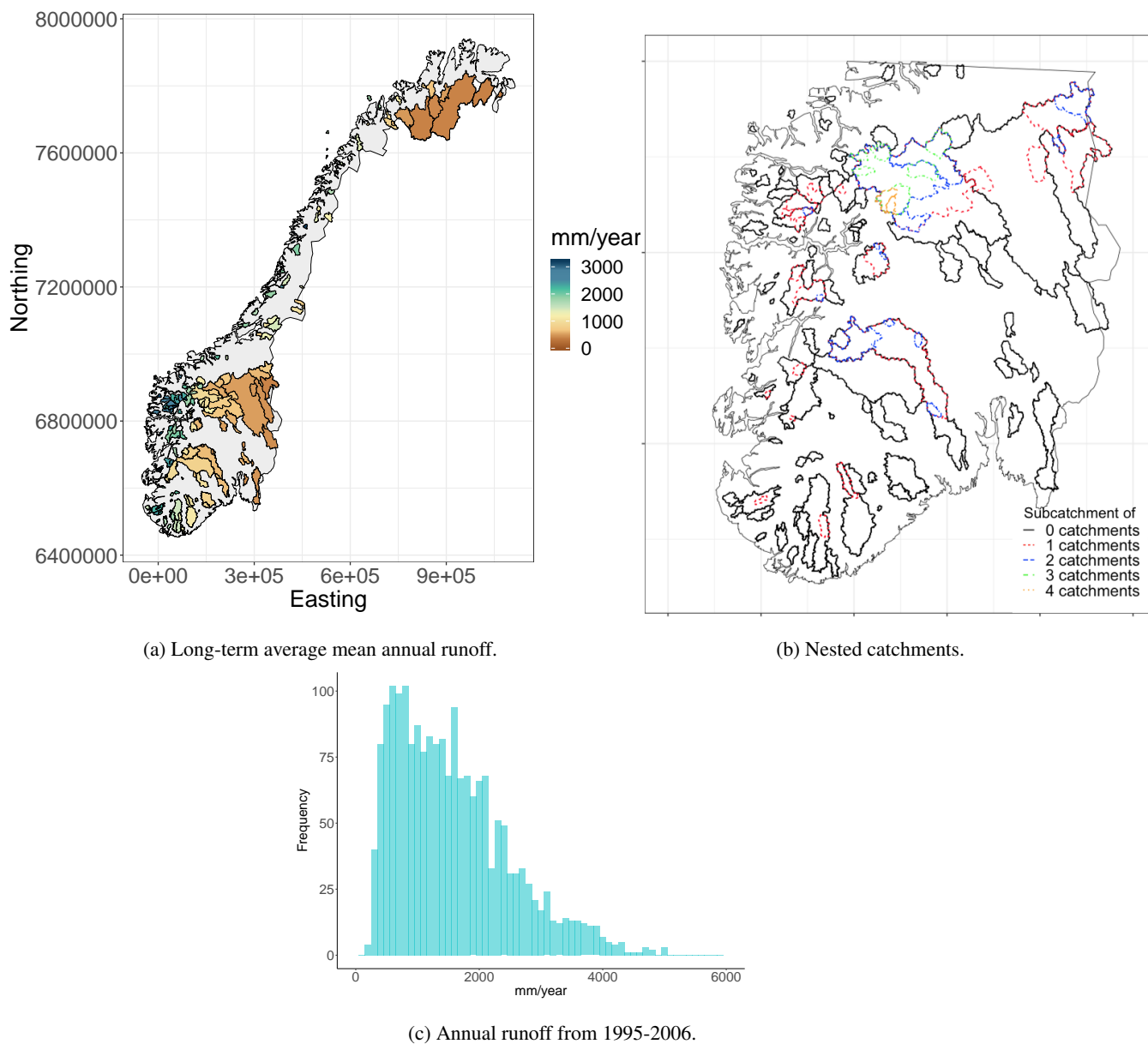
(a) Long-term average mean annual runoff.

(b) Nested catchments.

(c) Annual runoff from 1995-2006.

Figure 1. Long-term mean annual runoff (1996-2005) from 195 catchments in Norway (1a) and annual runoff observations from all 195 catchments and years (1c). Many of the catchments are nested, particularly in southern Norway as visualized in Figure 1b. In this figure, all colored catchments are subcatchments of at least one larger catchment, while the black catchments are not subcatchments of any larger catchment (but might contain 1-4 smaller catchments). In the visualization in Figure 1a, catchments with large areas are plotted behind catchments with smaller areas, and this is done throughout the article. The coordinate system used is EUREF89 - UTM33N (EPSG 25833). See Figure 6 for a closer image of the true annual runoff in southern Norway.
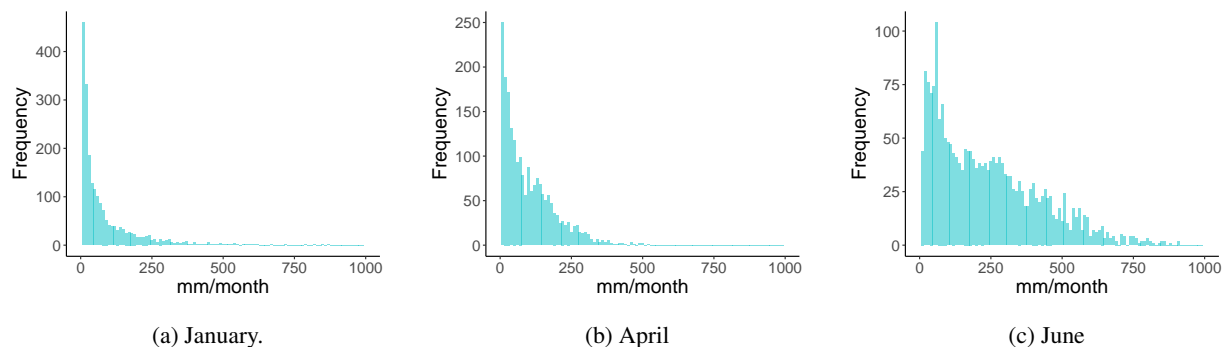
Figure 2. Monthly runoff data (1996-2005) from around 200 catchments in Norway for January, April and June.

for most catchments. 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

5   snow. In these regions the highest monthly runoff is observed for the snow melt season (May – June).

Monthly time series from the 7 selected catchments from Figure 3a are shown in Figure 4. We see that the year to year variability is higher, and the spatial pattern is less stable on a monthly scale compared to the annual scale (Figure 3b). In particular for months and locations where the temperature is close to zero degrees, we can expect unstable spatial patterns for monthly runoff. Monthly runoff data from Norway are thus suitable for evaluating the suggested method's performance on a

10 more unstable spatial pattern. In Figure 4 we see that June is the month with the most stable pattern, whereas January has the most unstable pattern. However, there is still a climatic trend that can be exploited by a model with the described properties.

As mentioned in the introduction, we also use our suggested interpolation framework to produce a runoff map for mean annual runoff in southern Norway for 1981-2010. For this purpose, we utilize all available data in this 30 year period, and there are data from 292 catchments for which only 89 of them have data from the whole period. The available observations for

15 1981-2010 are visualized later, in Section 5.5 (Figure 15).

## 3   Statistical framework for runoff interpolation

In Section 3.2 we present two Bayesian geostatistical models for runoff interpolation particularly suitable for sparse datasets containing several missing values and short records of runoff data. First, some statistical background is necessary.
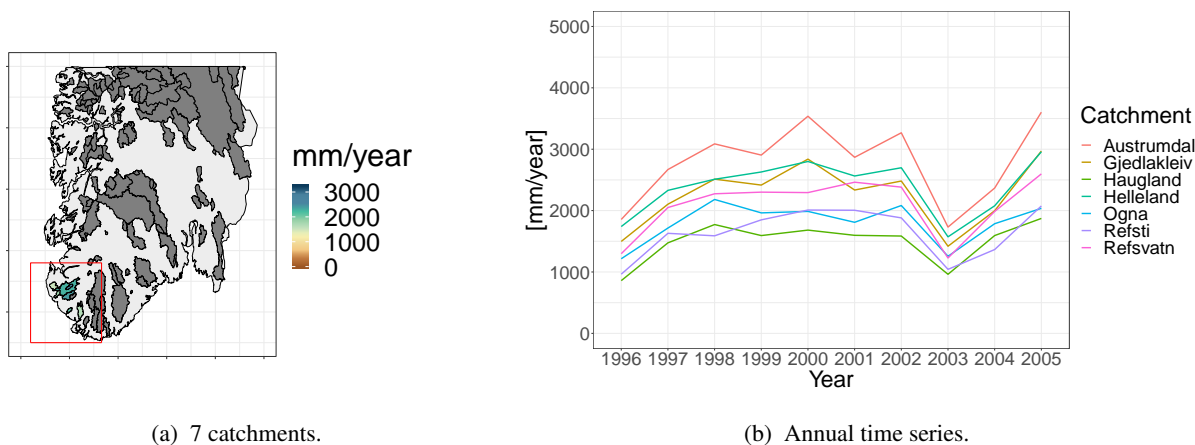
(a) 7 catchments.

(b) Annual time series.

Figure 3. Time series of annual runoff from 7 selected catchments in Western Norway. The 7 lines are almost parallel indicating that most of the spatial variability can be explained by climatic conditions.
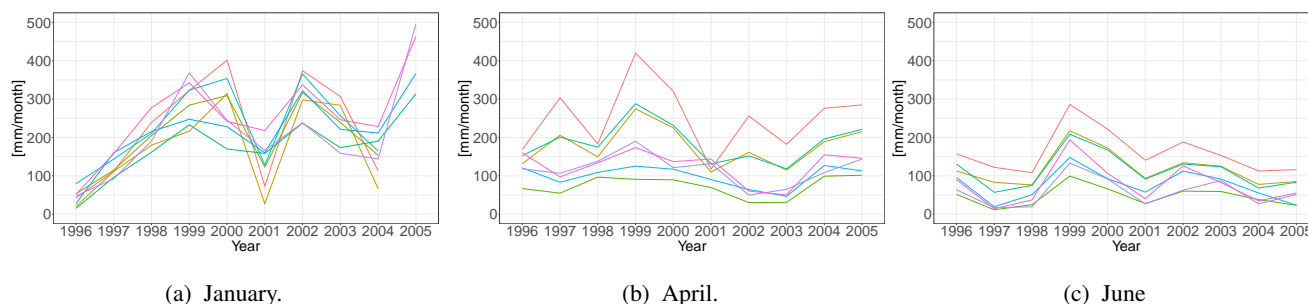
(a) January.

(b) April.

(c) June

Figure 4. Time series of monthly runoff for January, April and June for the 7 catchments in Figure 3a. The time series for January and April are less parallel than for June or for the annual runoff in Figure 3b, but there is still is a large correlation between the station.

3.1 Statistical methodology

3.1.1 Bayesian statistics and hierarchal modeling

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

Assuming x is a vector consisting of hydrological variables of interest, e.g the monthly rainfall for a specific year, the observation likelihood $\pi(y|x)$ expresses how the data y, e.g the observed monthly rainfall from some other years, are connected to x. In the classical frequentistic approach, the variables x are considered as unknown, but fixed. In the Bayesian approach
10 however, the variables x are considered to be a quantity whose variation can be described by a probability distribution (see

7

Hydrology and
Earth System
Sciences
Discussions
Open Access
EGU

e.g Casella and Berger (1990)). Prior to the analysis, this probability distribution is expressed through what is called a prior distribution $\pi(\boldsymbol{x})$ that is constructed based on expert knowledge about the variable of interest. The goal of the Bayesian analysis is to update this prior distribution by using data. By using Bayes' formula, the so-called posterior distribution of $\boldsymbol{x}$ is obtained:

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

5

Next, the marginal distribution $\pi(x_i|\boldsymbol{y})$ for $x_i \in \boldsymbol{x}$ can be integrated out, and a point prediction of x_i can be reported as 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 the process by thinking of things in a hierarchy of processes or distributions (Banerjee et al., 2004). E.g monthly rainfall $\boldsymbol{x}$ can be thought of as a process that depends on some parameters $\boldsymbol{\theta}$ that express the spatial correlation between the precipitation gauges. 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-stage 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 and monthly runoff.

3.1.2 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 $(x(\boldsymbol{u}_1), ..., x(\boldsymbol{u}_n))$ follows a multivariate normal distribution (Cressie, 1993), i.e $(x(\boldsymbol{u}_1), ..., x(\boldsymbol{u}_n)) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}$ is a vector of expected values and $\boldsymbol{\Sigma}$ is the covariance matrix. The covariance matrix $\boldsymbol{\Sigma}$ defines the dependency structures 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 provides information about how the covariance between the process at two locations decays with distance. The range is defined as the distance for 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 defined 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 $\nu > 0$, 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 where the covariance between two locations has dropped to 0.1. Using a Matérn GRF is

5   convenient for computational reasons because it enables us to apply the SPDE approach to spatial modeling from Lindgren et al. (2011) which is briefly described in Section 3.3.

3.1.3 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

10   represented as a random field, typically a Gaussian random field $x(\boldsymbol{u})$ with a known covariance structure given 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 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

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

15   where λ_i are interpolation weights. 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 value that both minimizes mean squared error, and that gives zero mean expected error (Cressie, 1993).

In order to minimize the mean squared error of the Kriging-predictor in Equation (3), the covariance function (or variogram) must be evaluated. The covariance function typically depends on the distance between the observations and the target locations,

20   such that observations collected close to the target location $\boldsymbol{u_0}$ are weighted more than observations further away. For streamflow related variables, the measure of distance is not given as the observed values often are connected to (catchment) areas, and not to single point locations in space. Catchments are also organized into subcatchments, and this should be taken into account when computing the Kriging weights. 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

25   violation of basic mass conservation laws.

The Top-Kriging approach suggested by Skøien et al. (2006) takes the nested structure of catchments into account when computing the distances and the Kriging-weights. It differs from other Kriging methods by weighting observations from a subcatchment more than an observation from a nearby non-overlapping catchment. Top-Kriging assumes linear aggregation. Thus, the method applies to variables that are mass conserved over nested catchments. However, it has been shown that Top-

30   Kriging also performs well for variables that are not mass conserved, like e.g the specific 100-year flood. Top-kriging is currently one of the leading methods for interpolation of hydrological variables (Viglione et al., 2013) and is therefore chosen as the method to compare our interpolation method with.

3.1.4 Methods for exploiting short records

The framework we suggest is both a framework for spatial interpolation and a framework for exploiting short records of runoff data. There are several ways to exploit short records of runoff data for which most of them are based on linear regression methods, utilizing correlation between catchments to improve the hydrological predictions and/or scale two time series relatively to

5 each other (Fiering, 1963; Laaha and Blöschl, 2005). In this article, we simply choose simple linear regression as the method of comparison.

Assuming 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

10 (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, $y_i = \beta x_i + \epsilon_i$ 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 is a coefficient that has to be estimated. The linear relationship between the two catchments is developed by estimating β by minimizing the sum of least squares, $\sum_{i=1}^{n}(y_i - \beta x_i)^2$. The linear relationship

15 can then be used to estimate the target variables $y_{n+1}, ..., y_{n+m}$ with uncertainty based on $x_{n+1}, ..., x_{n+m}$.

3.2 A geostatistical framework for exploiting short records by utilizing long-term spatial averages

In Section 3.2.1 - 3.2.6, we present the suggested Bayesian geostatistical framework. We first develop a framework for annual runoff, before we in Section 3.2.7 explain how the methodology also can be used for monthly runoff. We start by developing a three staged hierarchical model consisting of a process model, observation likelihood and prior distribution as described in

20 Section 3.1.1.

3.2.1 True annual runoff (process models)

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

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

25 $$\pi(\beta_c) \sim \mathcal{N}(0, 10000^2);$$

$$\pi(\beta_j | \sigma_\beta) \sim \mathcal{N}(0, \sigma_\beta^2)$$

$$\pi(c(\boldsymbol{u}) | \rho_c, \sigma_c) \sim \mathrm{GRF}(\rho_c, \sigma_c)$$

$$\pi(x_j(\boldsymbol{u}) | \rho_x, \sigma_x) \sim \mathrm{GRF}(\rho_x, \sigma_x)$$

where β_c is an intercept common for all years $j = 1, ..r$ that models the average runoff in Norway over time, while β_j is an

30 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 average of runoff, or the spatial variability caused by climatic conditions in Norway, while $x_j(\boldsymbol{u})$ is a year

specific spatial effect that models the spatial variability due to annual discrepancy from the climate. Both spatial effects are modeled as Gaussian random fields 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 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 $\mathcal{N}(0, 10000^2)$.

We suggest two alternative models for the true average annual runoff generated inside a catchment $\mathcal{A}$ in year j. 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 a catchment area, i.e

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

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 any point in the landscape (with some uncertainty). Thus, the areal model is a model particularly suitable for mass-conserved hydrological variables. It also gives a realistic representation of distances and thus also the correlation (Equation 2) between the catchments under study.

The second model is denoted **the centroid model**. For the centroid model, the true average annual runoff in catchment $\mathcal{A}$, 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 in Equation (4), and $\boldsymbol{u}_\mathcal{A}$ is the centroid of catchment $\mathcal{A}$. This alternative does not require preservation of water-balance and can be used for any environmental variable. Distance is measured between catchment centroids, such that this method is more similar to the traditional Kriging-methods described in Section 3.1.3. By skipping the integral and not requiring any constraints to be fulfilled, a large reduction in the computational cost is obtained compared to the areal model.

3.2.2 Observation likelihood

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

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

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 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 3.2.3.

3.2.3 Prior models

There are 6 model parameters $(\sigma_y, \rho_c, \sigma_c, \rho_x, \sigma_x, \sigma_\beta)$ in the third stage of the suggested hierarchical model for annual runoff.

5 As we apply the Bayesian framework, these have to be assigned with prior distributions, and we use knowledge based priors for most parameters.

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 m/year to 4000 m/years, and heteroscedastic errors can be

10 assumed. 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:

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

where y_{ij} is the observed runoff in catchment i in year j. For the variance 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. The

15 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 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

20 $$\text{Prob}(\sigma_y > 1500 \ \text{mm}) = 0.1. \tag{10}$$

Recall that it 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 approximately $(0.001, 10)\%$ of the observed value y_{ij}. This corresponds well to what we know about the measurement uncertainty for runoff in the study area.

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 $c(\boldsymbol{u})$,

25 we utilize the joint informative PC-prior suggested in Fuglstad et al. (2015). It is specified by the following probabilities and quantiles:

$$\text{Prob}(\rho_x < 20 \, \text{km}) = 0.1, \quad \text{Prob}(\sigma_x > 2000 \, \frac{\text{mm}}{\text{year}}) = 0.1,$$

$$\text{Prob}(\rho_c < 20 \, \text{km}) = 0.1, \quad \text{Prob}(\sigma_c > 2000 \, \frac{\text{mm}}{\text{year}}) = 0.1.$$

The percentages and quantiles are chosen based on knowledge about the spatial variability in the area of interest. It is reasonable

30 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 400 mm/year - 4200 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 this way detect hydrological stability or instability.

5   As specified in Section 3.2.1, 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 $P(\sigma_\beta > 10 \text{ m}) = 0.2$. With this prior, the prior 95% credible interval is approximately $(0.002, 40.5)$ m for the standard deviation σ_β of β_j.

3.2.4   Feasible computation of catchment runoff for the areal model

10   In the areal model in Equation (5), 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
15   subcatchment shares grid nodes with its mother catchment such that the water-balance can be preserved. In our analysis, we use a regular grid with 4 km spacing. Many of the Norwegian catchments are overlapping as we showed in Figure 1b, and in some areas each grid node is utilized in the discretization of as much as five catchments.

3.2.5   Hierarchical geostatistical model with short records

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
20   of the catchments $\mathcal{L}_{\mathcal{A}_i}$ for $i = 1,...,n$, the areal model in Section 3.2.1 - 3.2.4 is summarized as the following hierarchical geostatistical model:

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

$$+ 1 \cdot I\{\text{Observation } y_{ij} \text{ is missing}\}) \quad [\text{Observation likelihood}]$$

25    $$\pi(\boldsymbol{x}|\boldsymbol{\theta}) = \pi(c(\boldsymbol{u_1}),...,c(\boldsymbol{u_m})|\rho_c, \sigma_c) \cdot \pi(\beta_c)$$

$$\cdot \prod_{j=1}^r [\pi(x_j(\boldsymbol{u_1}),...,x_j(\boldsymbol{u_m})|\rho_x, \sigma_x) \cdot \pi(\beta_j|\sigma_\beta)] \quad [\text{Latent Model}]$$

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

Hydrology and
Earth System
Sciences
Discussions
Open Access

EGU

where $Q_j(\mathcal{A}_i)$ is the true annual runoff given by Equation (5), $I(\cdot)$ is a indicator function that is equal to 1 if its argument is true, and 0 otherwise allowing for missing data and short records of runoff, $\boldsymbol{y}$ 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. Finally, $\boldsymbol{\theta} = (\rho_x, \sigma_x, \rho_c, \sigma_c, \sigma_\beta)$. Together with

5 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 (5). 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.

3.2.6 Two model properties

10 There are two properties of the suggested interpolation scheme for runoff predictions that should be highlighted. These two properties makes the suggested models different from Top-Kriging and geostatistical interpolation methods that are typically used for hydrological applications.

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

15 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 average of runoff. If most of the spatial variability can be explained by long-term averages, 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). Thus, a short-record of runoff from an otherwise ungauged catchment will have a large impact also for the predictions in years without data through $c(\boldsymbol{u})$.

20 Figure 3 shows 7 time series of annual runoff from 7 randomly chosen catchments from Western Norway. The 7 lines are almost parallel, i.e the correlation between the annual runoff in these catchments is large indicating that the spatial variability is stable over years. This is captured by $c(\boldsymbol{u})$. By gathering one or a few observations of annual runoff from an otherwise ungauged catchment, $c(\boldsymbol{u})$ is updated and the model learns how the level of annual runoff in this catchment is relatively to longer time series from nearby catchments, i.e the nearby catchments work as donors. Existing methods that exploit short

25 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 (Roksvåg et al., 2019).

30 The second property we highlight only holds for the areal model, and is the model's ability to do more than smoothing: Runoff is in Equation (5) defined as an integral of point runoff such that the water-balance is preserved in any point in the landscape. This also has the beneficial 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 three catchments in Figure 5, where each black node represents one areal unit denoted Δ. The observed runoff is 1000 mm/year and 2000 mm/year in the two
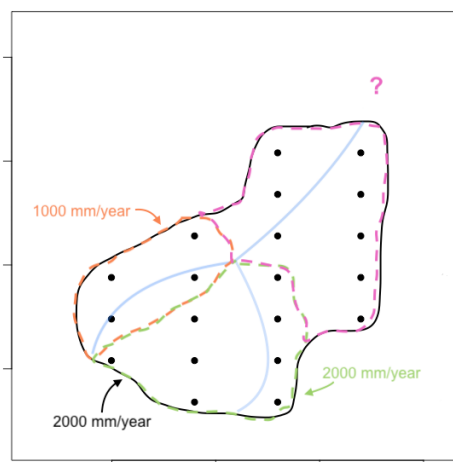
Figure 5. Conceptual figure of a river network. Assume that there are three observations of annual runoff: 1000 mm/year and 2000 mm/year in the sub-catchments (orange and green), and 2000 mm/year in the surrounding catchment (black). In order to fulfill the water balance, the predicted mean annual runoff in the remaining area (pink) must be 2500 mm/year with some uncertainty added (assuming each grid node represents 1 areal unit).

subcatchments, and 2000 mm/year in the surrounding larger catchment (with some uncertainty). In order to preserve the water-balance, the mean annual runoff x in the remaining area is given by the following equation:

$$1000 \text{mm/year} \cdot (4\Delta) + 2000 \text{mm/year} \cdot (8\Delta) + x \cdot \text{mm/year} \cdot (8\Delta) = 2000 \text{mm/year} \cdot (20\Delta) + \text{Uncertainty}$$

If the uncertainty is low, this results in a prediction around $x = 2500$ mm/year for the remaining part of the catchment.
5   Most Kriging-methods, involving the centroid model, do a weighting of the observations according to Equation (3) and would produce a prediction between 1000 m/year and 2000 m/year (Adamowski and Bocci, 2001; Merz and Blöschl, 2005; Skøien et al., 2006). The areal model suggested in this article is capable of producing predictions that are larger than any observed value without using any explanatory variables. However, this property also restricts these models to hydrological variables that are approximatively mass conserved over nested catchments. This is also one of the reasons for presenting the alternative
10   centroid model that can be used for any environmental spatial variable.

3.2.7   Extension to monthly runoff

In Section 3.2.1-3.2.6 we presented two models for annual runoff. However, the model can also be used to model other runoff signatures. More specifically, $Q_j(\mathcal{A}_i)$ in Equation (5) can for example represent the true total 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
15   the annual spatial discrepancy from the climate in January, and y_{ij} is the observed runoff in January for catchment $\mathcal{A}_i$ year j. In this article, we consider both annual predictions and monthly predictions for three chosen months (January, April and June), and for simplicity we use the same prior distributions for all time scales.

3.3 Inference

In order to make the models computational 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 of matrix operations on dense covariance matrices are required. The computational complexity is largest for the areal model, for which each

5 grid node can be regarded as a new target location. To solve the computational issues, 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 there exist efficient algorithms to perform

10 fast matrix operations on such matrices (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 have Bayesian models including a large number of parameters for which the marginal distribution 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. These challenges are met

15 by modeling runoff as a latent Gaussian model (LGM). That is, the latent part of the hierarchical model in 3.2.5 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 can be used for making Bayesian inference on LGMs and is a fast and approximative alternative to MCMC algorithms. The approach is based

20 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 described.

The R-package `r-inla` was utilized to make inference and predictions for the suggested models. This package provides a

25 user-friendly interface for applying INLA and the SPDE approach to spatial modeling without requiring that the user has deep knowledge about SPDEs. 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` while the supplementary material in Roksvåg (2016) explains how a climatic GRF like $c(\boldsymbol{u})$ can be introduced.

4 Model evaluation

4.1 Experimental set-up for cross-validation

30

To assess the predictive performance of our suggested models, we perform predictions of annual runoff, and monthly runoff for January, April and June for 1996-2005 for the test data described in Section 2. Recall that the test data consist of annual and

monthly data from around 200 fully gauged catchments located all across Norway. The three following methods are compared:

Top-kriging: Spatial interpolation with Top-Kriging. For Top-Kriging each year (1996-2005) is treated independently from other years. Short records on an annual (or monthly) scale don't have an impact on years without data. The default covariance

5   function (or variogram) in the R package `rtop` was used 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 3 with true annual runoff given by the areal model in Equation (5). That is, annual runoff is interpreted as the average point runoff over nested catchment areas. All years are

10   modeled simultaneously (1996-2005) such that short records of data might influence years without data.

Centroid model: Spatial interpolation with the model defined in Section 3 with true annual runoff given by Equation (6). That is, annual runoff is interpreted as a process linked to point locations in space (the catchment centroid), and not to catchment areas. All years are modeled simultaneously (1996-2005) such that short records of data might influence years without data.

15   The predictive performance of the three methods is evaluated by cross-validation. The around 200 catchments in Norway were divided into 20 groups or folds. 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 around 10 target catchments at once by using data from the remaining 190 catchments, and repeat the process for all 20 cross-validation groups. To evaluate and compare the three methods described above, we do the following three tests:

20

1) **UG (ungauged):** Assess the method's ability to predict runoff in 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 annual or monthly predictions for 1996-2005 are performed.

2) **PG (partially gauged):** Assess the method's ability to predict runoff in partially gauged catchments (denoted PG). Each of

25   the 10 target catchments in the cross-validation group is allowed to have one observation of annual or monthly 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 catchments in the other cross-validation groups. The short-record is drawn randomly from the ten years of observations available for each target catchment. We perform annual and monthly predictions for 10 partially gauged target catchments at once, for all 10 years under study (for which one of them is observed for each catchment), and repeat the process

30   for all 20 cross-validation groups.

3) **PG-N (partially gauged neighbors):** Assess the method's ability to exploit short records of data from neighboring catchments (denoted PG-N). The target catchments are treated as ungauged, while their three nearest neighboring catchments in terms of catchment centroid are treated as partially gauged. Nine out of ten of the observations from the neighboring catchments are removed from the observation likelihood. That is, the three neighboring catchments are allowed to have only one

35   observation of annual (or monthly) runoff. This observation is drawn randomly from the ten years of observations available. Full data series from the remaining catchments located further away are available as before. As we have around 10 catch-

ments in each cross-validation group, it means that they together have around 30 neighbors for which we remove 9 out of 10 observations. Consequently, there is a large reduction of available data in this test.

To make the results comparable, we use the same cross-validation groups for all experiments (UG, PG and PG-N) and methods (Top-kriging, areal model and centroid model). For the PG-case, we also compare our models to a method for exploiting

5 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 utilized as a donor catchment and only catchments outside the target catchment's cross-validation group can be considered. Two years of 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

10 regression model as described in Section 3.1.4. Next, the fitted model is used to predict runoff for the target catchment for 1996-2005 (where two of the years are observed).

4.2 Evaluation scores

As evaluation scores for the cross-validation we use the Root Mean Squared Error (RMSE) and the Continuous Rank Probability Score (CRPS). We compute the RMSE for catchment $\mathcal{A}_i$ as

15 $$\text{RMSE}_i = \sqrt{\frac{1}{r}\sum_{j=1}^{r}(y_{ij} - \hat{y}_{ij})^2},$$

where $\hat{y}_{ij}$ is the predicted annual runoff or predicted monthly runoff for catchment i year j, and y_{ij} is the corresponding observed value. The posterior mean is used as a point prediction for the areal and the centroid model. The average RMSE_i over all catchments $i = 1,..,n$ is used as summary score, and we denote it $\overline{\text{RMSE}}$.

The CRPS is defined as

20 $$\text{CRPS}(F_{ij}, y_{ij}) = \int_{-\infty}^{\infty}(F_{ij}(u) - 1\{y_{ij} \leq u\})^2 du,$$

where F_{ij} is the predictive cumulative distribution and y_{ij} is the actual runoff observation (Gneiting and Raftery, 2007). In this case, F_{ij} is a Gaussian distribution with mean equal to the predicted value and standard deviation equal to the standard deviation of the prediction. As for the RMSE, we denote the average CRPS over all catchments and years as $\overline{\text{CRPS}}$, while CRPS_i is the average CRPS in catchment i over r years. Both the CRPS and the RMSE are negatively oriented such that a low

25 score means an accurate prediction.

We also want to compare our results with other studies of annual and monthly 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. ANE is given by

$$\text{ANE} = \frac{|y_{ij} - \hat{y}_{ij}|}{y_{ij}}, \tag{12}$$

such that the absolute difference between the actual observation y_{ij} and corresponding prediction $\hat{y}_{ij}$ is normalized with respect to the magnitude of the observed runoff. An ANE close to zero corresponds to an accurate prediction. We use ANE_i to denote the average ANE in catchment i over r years. The squared correlation coefficient for catchment i is computed as

$$r_i^2 = (\mathrm{Cor}\{(y_{i1}, ..., y_{ir}), (\hat{y}_{i1}, ..., \hat{y}_{ir})\})^2, \tag{13}$$

where $\mathrm{Cor}(\cdot, \cdot)$ denotes the Pearson correlation, and it is taken between the observations y_{ij} from catchment i and the corresponding predictions $\hat{y}_{ij}$ for all years $j = 1, ..r$.

4.3 Mean annual runoff map for southern Norway

In addition to assessing the predictive performance of the suggested framework, we demonstrate our method on a real practical problem by constructing a map of mean annual runoff for southern Norway. For this purpose, all available data from 1981-2010 are used, i.e data from 292 catchments for which only 89 of them are fully gauged in this 30 year period. The data are used to estimate the mean annual runoff for around 1500 catchments, and to decrease the computational complexity of modeling 30 years of annual runoff simultaneously, we use the centroid model to do the calculations.

As the true underlying runoff field is unknown, the quality of the resulting runoff map cannot be quantified directly. However, we can compare our runoff map to alternative runoff maps, and for this purpose we also fit an alternative centroid model for which we omit the climatic GRF $c(\boldsymbol{u})$ in Equation (4). For this model, true annual runoff $q_j(\boldsymbol{u})$ at location $\boldsymbol{u}$ in year j is given by

$$q_j(\boldsymbol{u}) = \beta_c + \beta_j + x_j(\boldsymbol{u}); \qquad j = 1, ...r, \tag{14}$$

where the true annual runoff in catchment $\mathcal{A}$ is given by $q_j(\boldsymbol{u}_\mathcal{A})$ where $\boldsymbol{u}_\mathcal{A}$ is the catchment centroid as before (Equation (6)). We denote the above model "the simpler centroid model". The simpler centroid model treats each year of runoff more independently as it only has year specific effects. By comparing it to the full centroid model from Section 3.2, we can study the effect of modeling several years of runoff simultaneously.

5 Results

The models introduced in Section 3 (Top-Kriging, areal model, centroid model and linear regression) are fitted to the data described in Section 2. In Section 5.1 to 5.3 we present the results from the cross-validation and compare the method's performances for the three settings (UG, PG and PG-N) described in Section 4.1. In Section 5.4 we briefly compare the results with other studies of annual and monthly runoff, before we present the map for mean annual runoff for southern Norway in Section 5.5.

5.1 Predictions in ungauged catchments (UG)

For the ungauged case (UG), the target catchments are treated as totally ungauged for the ten study years 1996-2005, and annual and monthly predictions were performed by cross-validation. In Figure 7 the resulting average predicted *annual* runoff

Table 1. Predictive performance for predictions in ungauged catchments (UG).

Model	$\overline{\text{RMSE}}$ [mm]			$\overline{\text{CRPS}}$ [mm]			Coverage 95 %		
	TK	Areal	Centr.	TK	Areal	Centr.	TK	Areal	Centr.
Annual	**319**	363	360	**234**	266	263	0.93	0.93	0.91
January	**36**	38	37	**24**	25	**24**	0.92	0.92	0.87
April	**38**	39	**38**	**24**	25	25	0.92	0.88	0.83
June	**82**	87	94	**56**	60	66	0.90	0.92	0.81

Table 2. Predictive performance for predictions in partially gauged catchments (PG). Here, we also include results for linear regression (LR).

Model	$\overline{\text{RMSE}}$ [mm]				$\overline{\text{CRPS}}$ [mm]				Coverage 95 %			
	TK	Areal	Centr.	LR	TK	Areal	Centr.	LR	TK	Areal	Centr.	LR
Annual	302	184	203	**177**	209	**118**	133	256	0.94	0.95	0.93	0.96
January	34	**31**	32	86	21	**20**	21	145	0.90	0.90	0.87	0.94
April	36	**31**	31	46	22	**19**	20	83	0.86	0.86	0.83	0.98
June	79	**55**	64	87	51	**35**	44	180	0.90	0.90	0.81	0.96

Table 3. Predictive performance for predictions for ungauged catchments with partially gauged neighboring catchments (PG-N).

Model	$\overline{\text{RMSE}}$ [mm]			$\overline{\text{CRPS}}$ [mm]			Coverage 95 %		
	TK	Areal	Centr.	TK	Areal	Centr.	TK	Areal	Centr.
Annual	418	405	**366**	292	288	**265**	0.95	0.95	0.93
January	49	**47**	57	31	30	**29**	0.93	0.93	0.85
April	47	**44**	48	30	**28**	28	0.94	0.91	0.94
June	102	**100**	102	68	66	**65**	0.91	0.92	0.93

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 are all able to reproduce the true spatial pattern of annual runoff in Figure 6. The RMSE plots in Figure 7 also show that the three methods typically fails for the same catchments. Similar trends were seen for northern Norway and for monthly predictions.

5   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, with the areal model tending to give a slightly larger uncertainty. Top-Kriging and the areal model take the nestedness of catchments into account by implementing runoff data as areal referenced observations, providing a predictive standard deviation of runoff that depends on the size of the target catchment: Figure 7 shows that smaller catchments typically have a larger predictive uncertainty, which is reasonable. This is not the case for the centroid model. For
10 this 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 uncertainty provided by Top-Kriging and the areal method is thus more intuitive and realistic considering the process we are studying. The latter is also supported by the coverage percentages presented in Table 1. The coverage percentages 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.

5   Table 1 also presents the summary scores for the predictive performance for runoff predictions in ungauged catchments for all methods. In terms of RMSE and CRPS, Top-Kriging is a slightly better interpolation method than our two suggested methods. However, the boxplots in Figure 8 illustrate the distribution of RMSE_i for all catchments and years, and we see that the difference between Top-Kriging and the two other methods is quite low from a practical point of view. On a monthly scale the differences in RMSE_i are almost negligible.

5.2   Predictions in 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 present the results from the cross-validation for PG, 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 time-scales. The parameter estimates are presented in

15   Table 4 for annual and monthly predictions, for the areal and the centroid model. The parameters in this table are important, as they indicate how much of the spatial variability that lies in the climatic GRF relatively to the annual GRF. In particular, if σ_c dominates over σ_x, it suggests hydrological stability and we can expect short records to have a large impact on the predictive performance for the target catchment.

The estimates in Table 4 show that the hydrological stability is largest for April, June and for annual data. Particularly

20   for June and for the annual data, for which 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. The latter reflects a strong climatic trend where $c(\boldsymbol{u})$ dominates over $x_j(\boldsymbol{u})$. For January, the spatial variability is less stable from one year to another, with posterior modes of σ_c and σ_x that are both around 80 mm/month for the areal model, and with σ_x approximately twice as large as σ_c for the centroid model. Consequently, for annual predictions and for monthly predictions for June, short records should have a large impact on the predictive performance, while

25   for January and April, year specific-effects explain a larger part of the spatial variability, and the gain of including short records is expected to be lower.

The conclusions drawn for the parameter values, are confirmed by comparing the RMSE and CRPS for the areal and the centroid model in Table 2, to the RMSE and CRPS obtained for the ungauged case in Table 1. For all time-scales, the RMSE and CRPS for our two models are reduced in Table 2 compared to Table 1, although the reduction in the RMSE and CRPS is

30   low for January and April as expected from the parameters. For 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 40 %). The results holds for both the areal and centroid model, but the areal method seems to be somewhat better than the centroid model in terms of exploiting short records of data from the target catchment. This is
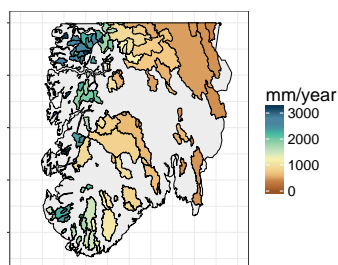
Hydrology and
Earth System
Sciences
Discussions

Open Access

EGU

Figure 6. Mean annual observations (1996-2005) for southern Norway.

Posterior mean.
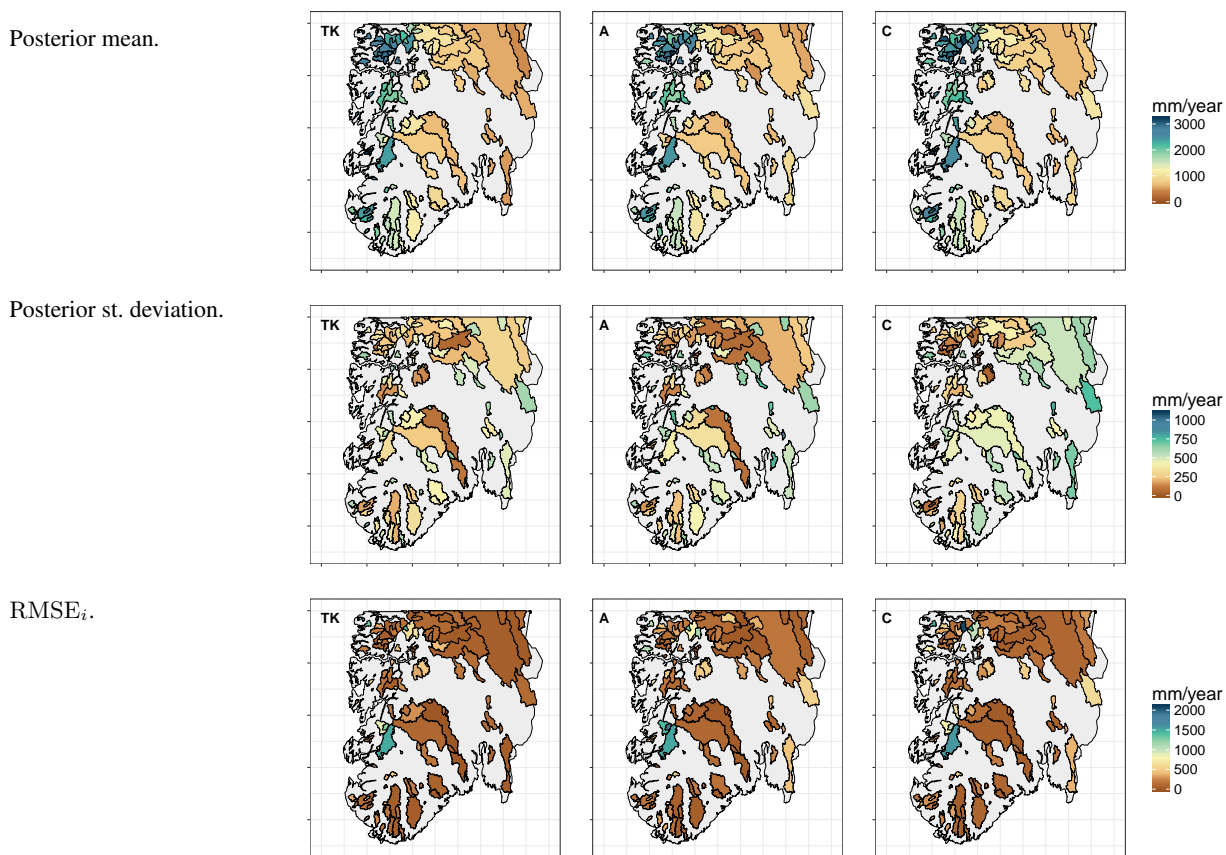
Posterior st. deviation.

RMSE_i.

Figure 7. Average posterior mean (upper), average posterior standard deviation (middle) and average RMSE for each catchment (lower) for annual predictions in southern Norway for the three interpolation methods (Top-Kriging, Areal and centroid) when the target catchments are treated as ungauged (UG). The methods give similar results for other time scales and for northern Norway.
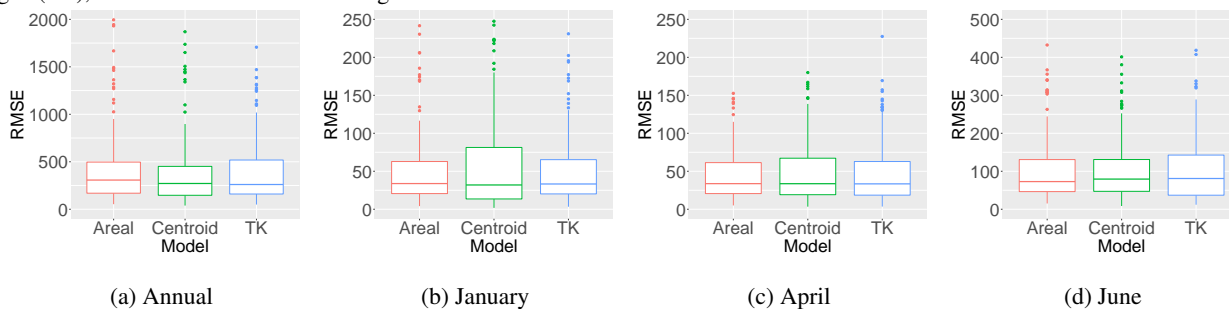
(a) Annual　　　(b) January　　　(c) April　　　(d) June
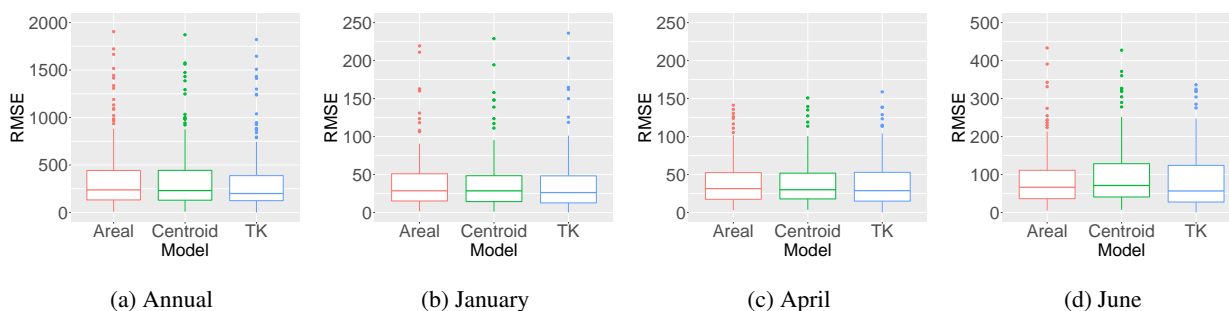
Figure 8. Distribution of $RMSE_i$ for predictions for all catchments and years (1996-2005) when the target catchments are treated as ungauged (UG) in the cross-validation.

(a) Annual　　　(b) January　　　(c) April　　　(d) June
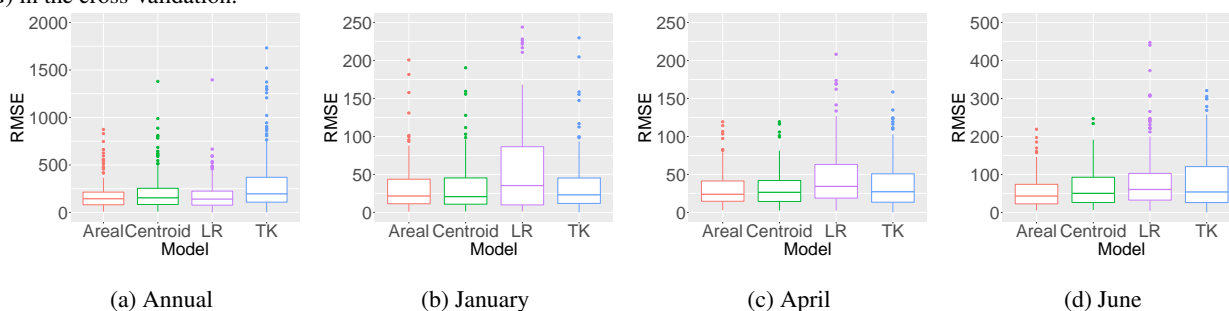
Figure 9. Distribution of $RMSE_i$ for predictions for all catchments and years (1996-2005) when the target catchments are treated as partially gauged (PG), i.e a short-record from the target catchment is included in the observation likelihood in the cross-validation.

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

Figure 10. Distribution of $RMSE_i$ for predictions for all catchments and years (1996-2005) when the target catchments are treated as ungauged and the three nearest neighboring catchments are treated as partially gauged (PG-N) in the cross-validation.

Hydrology and
Earth System
Sciences
Discussions

Open Access

EGU

Table 4. 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 slightly depending on which experiment and group in the cross-validation we consider, and the values given here are the mean over all folds and experiments (UG, PG and PG-N). The spatial effect that dominates (annual or climatic) is marked in bold.

Areal model	ρ_c [km]	ρ_x [km]	σ_c [mm]	σ_x [mm]
Annual	49 (6)	513 (74)	**902** (53)	270 (24)
January	17 (4)	285 (32)	80 (8)	**87** (4)
April	63 (9)	222 (40)	**74** (5)	48 (3)
June	35 (3)	182 (25)	**198** (10)	79 (4)

Centr. model	ρ_c [km]	ρ_x [km]	σ_c [mm]	σ_x [mm]
Annual	86 (11)	811 (115)	**697** (50)	271 (27)
January	57 (12)	495 (66)	49 (4)	**94** (9)
April	124 (18)	429 (70)	**67** (6)	52 (5)
June	64 (8)	364 (53)	**147** (10)	72 (6)

again related to the parameter estimates in Table 4, where we see that σ_c dominates more over σ_x in the areal model than in the centroid model.

While our suggested methods model several years of data simultaneously, Top-kriging treats each year of data independently. A reduction in RMSE and CRPS is only seen for the specific year with extra data. Thus, we only obtain a small reduction in

5   the RMSE and CRPS for the partially gauged case (Table 2) compared to the ungauged case (Table 1) for Top-Kriging. The evaluation scores in Table 2 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 stability. For January and April the three models are more similar.

For the PG case, we also compare the areal and the centroid model to simple linear regression. According to Table 2 and

10   Figure 9 linear regression performs quite well on an annual scale, which is the time-scale with most hydrological stability (according to σ_c and σ_x). Linear regression actually has the lowest RMSE of all four methods for annual predictions. However, recall that a short-record of length 2 from the target catchment is needed in order to use this method, while our areal model performs approximately equally well with a short-record of length 1 (and observations from other neighboring catchments). Table 2 also indicates that linear regression gives more unstable uncertainty estimates with a large CRPS for both annual and

15   monthly predictions. This can be explained by the fact that the variance of the measurement uncertainty σ^2 is estimated based on only two pairs of observations, i.e two observations from one donor catchment and one target catchment. For monthly predictions, linear regression is outperformed by the three other methods in terms of RMSE and CRPS (Table 2). We conclude that linear regression requires a quite high correlation between the target catchment and the donor catchment in order to produce better predictions than the geostatistical methods. Even for June which is a month with large hydrological stability,
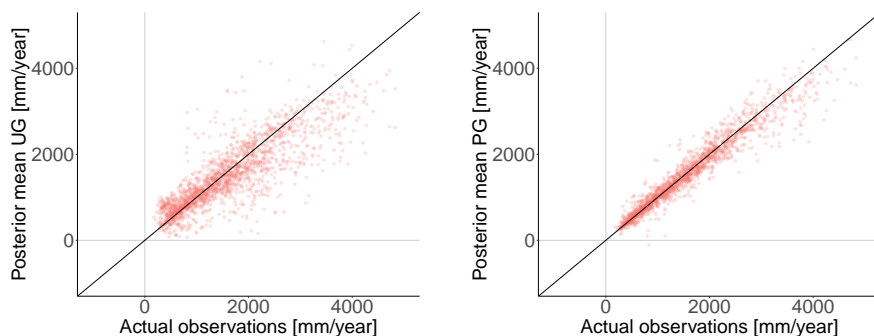
20   linear regression is outperformed.

Figure 11. Actual annual observations compared to the predictions for the ungauged case (UG, left) and the partially gauged case (PG, right) for the areal model. The straight line represents a perfect correspondence between prediction and actual observation.
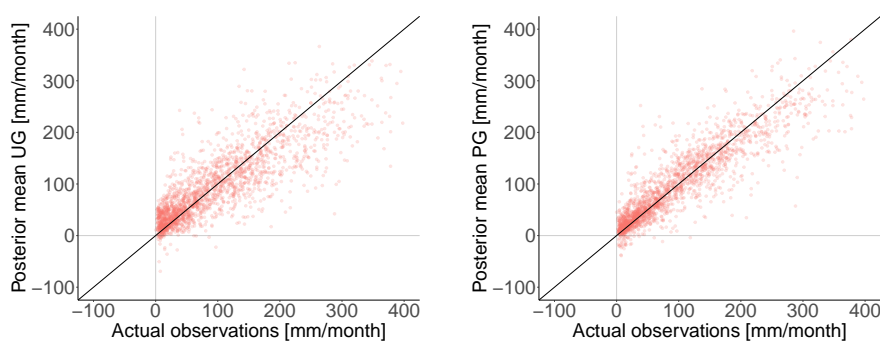

Figure 12. Actual observations from April compared to the predictions for April for the ungauged case (UG, left) and the partially gauged case (PG, right) for the areal model. The straight line represents a perfect correspondence between prediction and actual observation.

To illustrate the possible gain of including short records of data from the target catchment, we include some scatter plots that compare the predicted values produced by the areal model to the actual observations of runoff (Figure 11 and Figure 12). For the annual predictions in Figure 11, we see a large improvement in the predictions for the partially gauged case (PG) compared to the ungauged case (UG): While UG has some problems with predicting large values of runoff, the predictions for PG are

5   considerably more concentrated around the straight line that indicates a perfect fit. There are similar results for June, whereas the difference between the ungauged and partially gauged case is not that prominent for April (see Figure 12) and January.

In Figure 7 we saw that all three interpolation methods were able to reproduce the true spatial pattern of annual runoff when performing predictions in ungauged catchments (UG). However, all three methods had trouble identifying some of the catchments. These are typically wet catchment surrounded by dry catchments, or dry catchment surrounded by wet catchments.

10   Such catchments are difficult to predict for pure interpolation methods like the ones described in this article. Figure 13 shows the impact of including a short-record of length one for these problematic 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
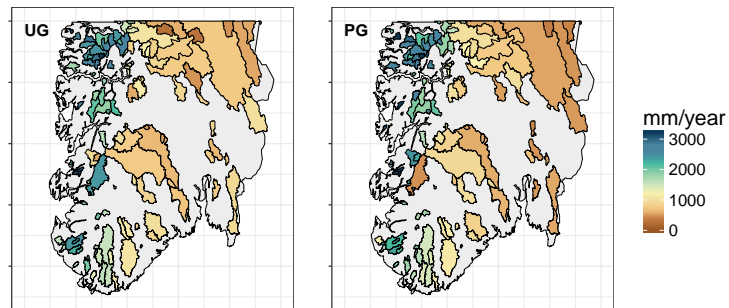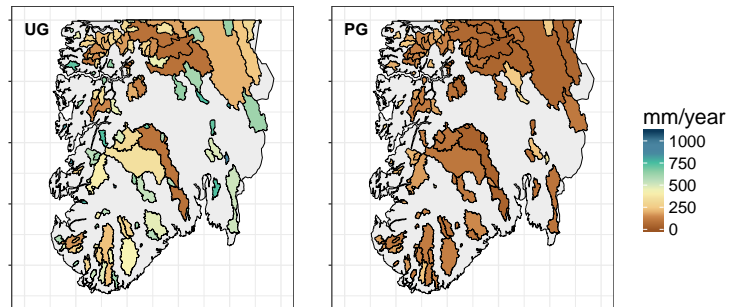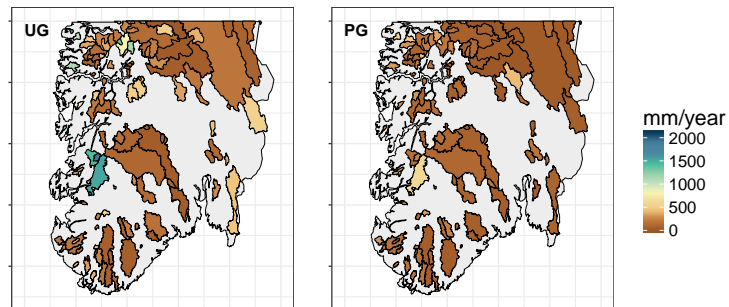
Figure 13. Average posterior mean (upper), average posterior standard deviation (middle) and RMSE_i (lower) for annual predictions performed by the areal model for the ungauged case (left) and the partially gauged case (right).

reduction in the RMSE, and a (realistic) reduction of the posterior standard deviation. There are similar findings for predictions in June. Figure 7 and 13 illustrate how a short-record of length one or larger can lead to a considerable reduction in the RMSE if the spatial pattern is stable over years, and the importance of having a model that can exploit this property.

### 5.3 Exploiting short records of data from neighbors (PG-N)

5   Next, we assess the method's ability to exploit short records of data from nearby catchments (PG-N), i.e the targets catchment are treated as ungauged, while three of the closest neighboring catchments are allowed to have only one observation of monthly or annual runoff. Table 3 and the boxplots in Figure 10 summarize the results from the cross-validation. For annual predictions, the centroid model gives the most accurate results with the lowest RMSE and CRPS. The centroid model outperforms the areal model here, possibly because the centroid model provides the largest spatial range (ρ_c=86 km for the centroid model compared

10   to ρ_c=49 km for the areal model). Thus, observations from nearby catchments have a larger impact on the target catchment for the centroid model than for the areal model for this case study.

   On a monthly scale, the three methods perform approximately equally good. One possible explanation is that the spatial range is not large enough for June, which is the month with most hydrological stability, for the target catchments to be considerably affected by short records of data from neighbors. The annual predictions for the areal model are also only slightly better than

15   the predictions provided by Top-Kriging. However, recall that for the ungauged case (UG) Top-Kriging outperformed the centroid and the areal method for all time scales (Table 1), while for the PG-N case the differences between the three methods are almost reduced to zero (Table 3). The centroid model and the areal model are even better than Top-Kriging on an annual scale in Table 3. The results from UG and PG-N in Table 1 and 3, indicate that exploiting long-term spatial averages in the interpolation scheme as in our methods, can be just as important as finding a process-based way of determining the Kriging

20   weights which is the idea behind Top-Kriging.

5.4 Comparison with other studies

We now compare our results to other comparable studies. In the chapter about annual runoff by McMahon et al. (2013) in Blöschl et al. (2013), 34 studies of predictions of mean annual runoff from around the world are compared in terms of the absolute normalized error (ANE) from Equation (12) and the squared correlation coefficient r^2 from Equation (13). In regions

25   like Norway for which the potential evapotranspiration is less than 40 % of the mean annual precipitation, a median ANE around 0.25 is a typical result (Figure 5.27 in McMahon et al. (2013)). The ANE obtained for annual predictions produced by our suggested areal model is thus similar to the ANE obtained by other comparable studies, with median ANE around 0.25 as shown in Figure 14a. The median ANE is as low as 0.10 for the PG case for annual predictions.

   Furthermore, McMahon et al. (2013) report an r^2 between 0.60 and 0.99 for studies based on cross-validation with ∼150

30   involved catchments, or for methods based on spatial proximity like our suggested models (Figure 5.25 and Figure 5.26 in McMahon et al. (2013)). The r^2 for our areal model is presented in Figure 14b, and we see that it is between 0.75-0.90 for the UG and PG case for annual predictions, i.e within the range of values obtained by other comparable studies.

27
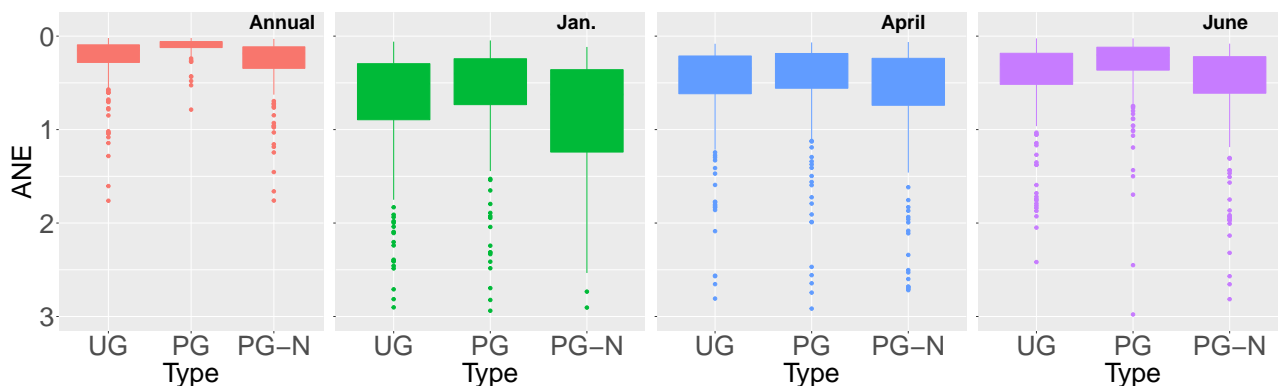
(a) Absolute normalized error (ANE_i) for each catchment for the areal model.
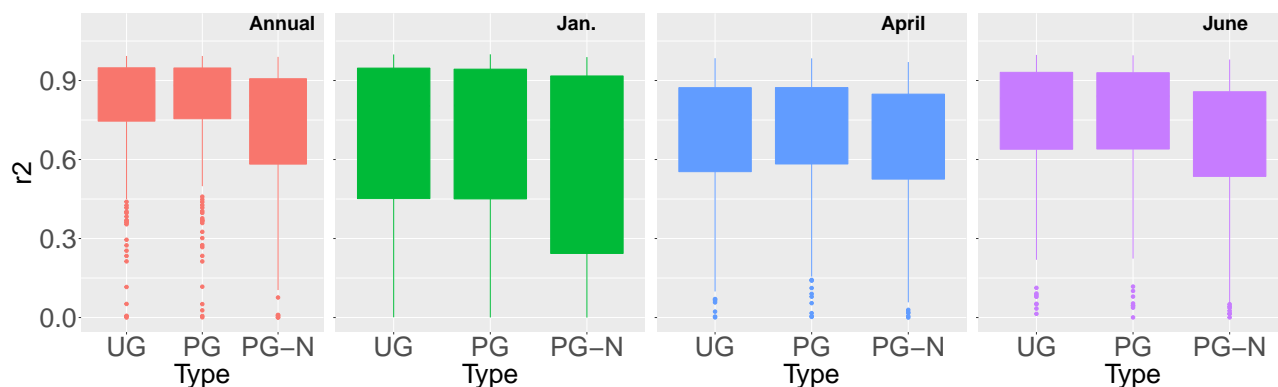
(b) Squared correlation coefficient r_i^2 for each catchment for the areal model.

Figure 14. Absolute normalized error (ANE_i) and squared correlation coefficient (r_i^2) for all catchments for the areal model. Cross-validation is performed on 200 Norwegian catchments as described in Section 4 for the three cases UG (ungauged catchments), PG (partially gauged catchments) and PG-N (partially gauged neighboring catchments).

In Blöschl et al. (2013) there is also a comparative study of seasonal runoff, but here the ANE and r^2 are reported for seasonal runoff as a whole, not for individual months like in this article. However, we include ANE and r^2 for monthly predictions in Figure 14 for completeness.

5.5 Mean annual runoff map for southern Norway

5 Finally, we demonstrate our framework for a real practical case by presenting two mean annual runoff maps for southern Norway: One that is constructed by the full centroid model, and one that is constructed by a simpler centroid model from Equation (14) for which the climatic GRF $c(\boldsymbol{u})$ is omitted. The resulting maps are shown in Figure 15. We see that the runoff field produced by the simpler centroid model is smoother than the runoff map provided by the full centroid model. Thus, the full model seems to be better at capturing rapid spatial variations. This is in particular visible in the western parts of Norway
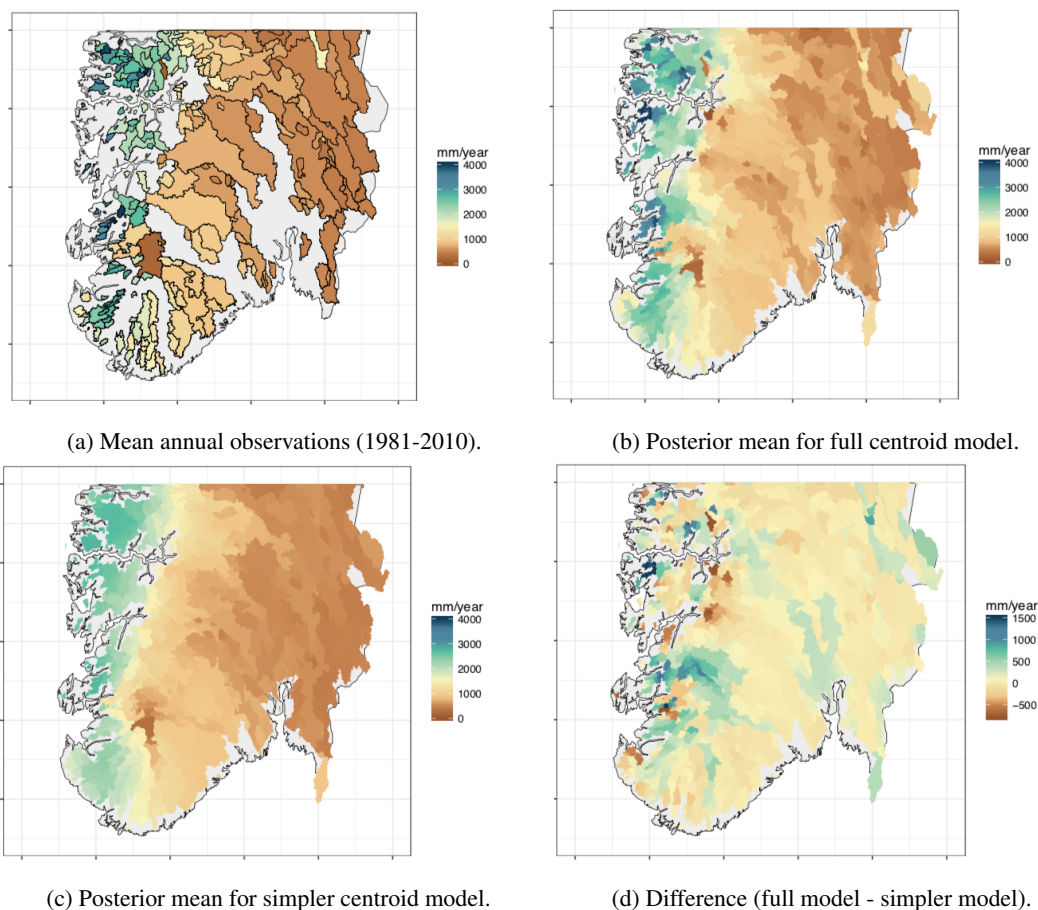
(a) Mean annual observations (1981-2010).

(b) Posterior mean for full centroid model.

(c) Posterior mean for simpler centroid model.

(d) Difference (full model - simpler model).

Figure 15. Average posterior mean annual runoff for all catchments in southern Norway (1981-2010) computed based on data from 292 catchments for which only 89 of them are fully gauged in the time period of interest.

close to the coast where there are several small catchments with large values of mean annual runoff. Some of these catchments deviate from their closest neighbors and don't fit into the underlying statistical model, but the full centroid model is able to account for this if a short-record is available here. As we can see, this holds both for catchments that are wetter or dryer than its surroundings. In the eastern parts of Norway (located in the rain shadow), the difference between the full centroid model and the simpler model is smaller, as we can see in Figure 15d. The weather conditions are more stable in space with a larger spatial range. Performing runoff predictions here are in general easier, and similar results for the two models indicate that short-records of runoff have less impact on the predictive performance in these areas.

6 Discussion

In this article we have presented two geostatistical Bayesian models particularly suitable for hydrological datasets that include missing values and short records of data. The models provide a framework for simultaneously modeling several years of runoff, and consist of one year specific GRF, and one climatic GRF that is common for all years under study. The climatic GRF detects hydrological stability and captures the long-term spatial variability in the area of interest. This information is used across years and allows the model to fully exploit data from partially gauged catchments.

One of our goals was to assess the predictive performance of our suggested framework, and compare this to the predictive performance of Top-Kriging. To quantify the predictive performance, we predicted monthly and annual runoff for around 200 catchments in Norway by cross-validation. The results confirmed that when the spatial variability is high, but stable over years, there is much predictive power to gain from utilizing a model with a climatic spatial field. For the Norwegian dataset, our two suggested models clearly outperformed Top-Kriging when the target catchments were treated as partially gauged (PG) with one observation from the target catchment included in the likelihood (while 10 years of runoff were predicted). A reduction in RMSE of around 50 % was found for annual runoff predictions in partially gauged catchments compared to predictions in totally ungauged catchments. One of our aims was also to assess the value of including short records of runoff, and by this we have shown that it can be large. Furthermore, there is also valuable information stored in short records of data from neighboring catchments (PG-N), but here the differences between Top-Kriging and our two suggested methods were smaller. When the target catchments were treated as ungauged (UG) and the neighboring catchments were fully gauged, Top-Kriging provided the best predictions. This indicates that Top-Kriging is the best method for pure interpolation of stream flow related variables. However, the results from the other experiments show that exploiting hydrological stability and long-term trends (as in our models) can be just as important as determining the Kriging weights in a process-based manner (as in Top-Kriging) if the dataset available is sparse.

Compared to simple linear regression with the closest catchment applied as donor catchment, our suggested methods were better at exploiting short-records from the target catchment in terms of RMSE and CRPS, except when the spatial pattern was particularly stable over years (annual predictions in Norway). For annual predictions linear regression performed slightly better than our (areal) model. However, our (areal) model provides a more realistic distribution of the measurement uncertainty, it can be applied for short-records of length 1 (linear regression requires length larger than 1), and represents an objective framework where we avoid the challenge of choosing a suitable donor catchment.

In addition to assessing the predictive performance of the suggested framework, our aim was to illustrate the benefits of including a climatic trend in the model for a real practical problem by constructing a runoff map for southern Norway. We constructed one map with our suggested framework, and one map based on a simpler model that only included annual GRFs that treat each year of data more independently. The resulting maps showed that the model that included a climatic GRF was better at capturing rapid changes of annual runoff than the simpler spatial model.

In order to compare our method to other studies, we computed the absolute normalized error (ANE) and the squared correlation coefficient r^2. The results showed that our methodology gave an ANE and r^2 within the range of values obtained

by other experiments. Furthermore, the resulting ANE and r^2 can also be used to better understand how our methodology works in practice. If we compare the results when performing predictions for ungauged catchments (UG) and the results when performing predictions in partially gauged catchments (PG), we see that the difference between UG and PG in terms of r^2 in Figure 14b is almost negligible, while the difference in terms of ANE in Figure 14a is large. Recall that r^2 is a measure of

5 correlation between the predictions and the actual observations, while ANE is a measure of deviance between the predictions and the observations. For the UG and PG case, we use the same set of observations in the likelihood for cross-validation, except for one extra observation from the target catchment for PG compared to UG. The strength of our suggested framework is that including an extra observation from the target catchment typically increases or reduces the level of predicted runoff in the target catchment relatively to its neighbors due to the climatic GRF. The correlation between the target catchment and the

10 neighboring catchments remains approximately unchanged. That is, including a short record from the target catchment does not have a large impact on r^2, or the correlation between the target catchment and its neighbors, but it often has a large impact on scores like RMSE or ANE because the level of annual or monthly runoff is changed. This way, the climatic component provides a methodology for calibration for the target catchment when a short-record is available.

In Section 5.2 we saw that it was a clear relationship between the marginal variances of the two spatial fields, and the

15 decrease in RMSE when a short record of length 1 was included in the likelihood. This represents another appealing property of the suggested model: Hydrological stability can be indicated directly from the parameters. If the marginal variance of the climatic GRF σ_c dominates over the marginal variance of the year specific GRF σ_x, this suggests that the spatial variability is stable and that short records of runoff have a large impact on the resulting predictions. This can be valuable information for decision-makers when deciding whether or not new observations should be gathered from an ungauged catchment, e.g when

20 planning a building project or the construction of a power station. However, to exactly quantify the expected gain of gathering a new stream-flow observation from an ungauged catchment, 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.

In this study, we proposed one model that ensures approximate mass conservation of runoff over nested catchments (areal

25 model), and one model for variables that are not mass conserved (centroid model). Both models provided accurate predictions of annual runoff, but the areal model gave a better interpretation of the predictive uncertainty and was also somewhat better in terms of exploiting short-records of runoff from the target catchment for the Norwegian dataset. As the areal model provides a more correct distribution of uncertainty over nested catchments, it might outperform the centroid model also for variables that are not mass-conserved. Top-Kriging performs well for such variables although it is a method that original was constructed

30 for mass-conserved variables (Skøien et al., 2006). There are also reasons to believe that the areal model will outperform the centroid model in areas with less data: In areas with few stream-flow observations, a more physical interpretation of runoff is probably more important. Here the areal model has some benefits as it constrains runoff over nested catchments, and is able to produce higher predicted values than any of the observations in the dataset. However, the centroid model is still useful, first and foremost due to its computational benefits. It can also easily be used for interpolation of other environmental variables such as

35 precipitation or temperature.

Like all methods, our suggested interpolation schemes have some shortcomings. For the pure interpolation case with fully gauged neighbor stations (UG), we saw that Top-Kriging performed better than our two methods. It is difficult to point out why Top-Kriging performs better, and we did not find a pattern for which catchments Top-Kriging performed best (mean elevation, location and the magnitude of the observations were investigated). One contributing reason for the difference in

5 predictive performance can be explained by the choice of covariance function: Our methods are restricted to applying the Matérn covariance function for computational reasons, while a multiplication of a modified exponential and fractal variogram model was used for Top-Kriging. Utilizing a Gaussian covariance function in the Top-Kriging method was tried, but gave poorer results. The Gaussian covariance function typically produced a spatial field that was too smooth, and this can be a problem for the Matérn covariance function too.

10 In this article, we proposed two models for runoff that are Gaussian. That is, there is nothing in the model that prevents it from predicting negative runoff. The 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 GRF can be negative in some areas. This is a fully valid result because the annual GRF and the other model components still ensure positive predictions for most catchments and years. However, it can become a problem if we are unlucky and the year

15 specific GRF lacks information from some of the partially gauged catchments for a particular year, such that the annual GRF doesn't make up for the negative climatic GRF in these areas.

In the areal model, negative values also appear as a consequence of requiring preservation of water-balance. Earlier, we described how the areal model is able to predict values that are larger than any of the observations, but it could also go the other way around and provide small or negative predictions. The latter happens particularly if there are inconsistent or poor data over

20 nested catchments. Also note that it is important that the discretization of the study area is fine enough to capture rapid changes in runoff over nested catchments to avoid unphysical results for the areal model. In our study, some negative values were produced for the annual and monthly predictions as we can see in Figure 11 and Figure 12. However, this is not common and happened for only 1.2 % of the predictions presented in this paper. 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

25 should be remarked, but are hard to fully avoid.

7 Conclusions

We have presented a new geostatisitcal framework for estimating flow indices by modeling several years of annual or monthly runoff data simultaneously by utilizing one (climatic) GRF that is common for all years under study, and one (annual) GRF that is year specific. By this, we obtain a framework that is particularly suitable for runoff interpolation when the available data

30 originate from a mixture of gauged and partially gauged catchments, and that can be used to estimate runoff at ungauged and partially gauged locations. The results from the case study of annual and monthly runoff in Norway showed that our modeling framework is well suited for exploiting the information stored in short records of runoff data, and for some catchments in the dataset the reduction in RMSE for annual predictions was as large as 50 % when a short record of length 1 was available. This

illustrates the importance of having a model that is able to transfer information across years, which is our method's main benefit compared to e.g Top-Kriging. By constructing a runoff map for annual runoff in southern Norway we also demonstrated that a model with the described properties is considerably better at capturing rapid changes of runoff than a similar model that treats each year of runoff data independently.

5   *Author contributions.* Thea Roksvåg: Main author, main responsible for writing and wrote the majority of the paper. Came up with initial ideas for experimental design. Did the implementation, carried out the analysis and made figures.

Ingelin Steinsland: Contributed to discussion throughout the work, around ideas, analysis and discussion. Suggested ideas for experimental set-up and commented on the manuscript structure and content.

Kolbjørn Engeland: Provided the hydrological data. Contributed to discussion, particularly around the hydrological context and questions
10 related to the data. Contributed to the writing of Section 1 and Section 2, and commented on the structure and content of the rest of the paper.

Competing interests. No competing interests are present.

Code and data availability. All data and code used in this study are available from the authors on request.

Hydrology and
Earth System
Sciences
Discussions

Open Access

EGU

References

Adamowski, K. and Bocci, C.: Geostatistical regional trend detection in river flow data, Hydrological Processes, 15, 3331–3341, 2001.

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

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

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

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

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

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

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H.: Interpretable Priors for Hyperparameters for Gaussian Random Fields, 2015.

Førland, E. J.: Nedbørens høydeavhengighet, Klima, 1979.

Gneiting, T. and Raftery, A. E.: Strictly Proper Scoring Rules, Prediction, and Estimation, Journal of the American Statistical Association,

15   102, 359–378, 2007.

Gottschalk, L.: Correlation and covariance of runoff, Stochastic Hydrology and Hydraulics, 7, 85–101, 1993.

Gottschalk, L., Jensen, J. L., Lundquist, D., Solantie, R., and Tollan, A.: Hydrologic Regions in the Nordic Countries, Hydrology Research, p. 273–286, 1979.

Guttorp, P. and Gneiting, T.: Studies in the history of probability and statistics XLIX On the Matérn correlation family, Biometrika, 93,

20   989–995, 2006.

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

Laaha, G. and Blöschl, G.: Low flow estimates from short stream flow records—a comparison of methods, Journal of Hydrology, 306, 264 – 286, 2005.

25 Lindgren, F., Rue, H., and Lindström, J.: 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.

McMahon, T., Laaha, G., Parajka, J., Peel, M., Savenije, H., Sivapalan, M., Szolgay, J., Thompson, S., Viglione, A., Woods, R., and Yang, D.: Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales., Camebridge University press, 2013.

Merz, R. and Blöschl, G.: Flood frequency regionalisation - spatial proximity vs. catchment attributes, Journal of Hydrology, 302, 283 – 306,

30   2005.

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

Roksvåg, T.: A Bayesian Model for Area and Point Predictions - A case study of Predictions of Annual Precipitation and Runoff in the Voss Area, Master thesis, Norwegian University of Science and Technology, 2016.

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

Hydrology and
Earth System
Sciences
Discussions

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

Sauquet, E., Gottschalk, L., and Lebois, E.: Mapping average annual runoff: A hierarchical approach applying a stochastic interpolation scheme, Hydrological Sciences Journal, 45, 799–815, 2000.

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

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

Skøien, J. O., Blöschl, G., and Western, A. W.: Characteristic space scales and timescales in hydrology, Water Resources Research, 39, 2003.

10 Stohl, A., Forster, C., and Sodemann, H.: 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, 2008.

Sælthun, N., Tveito, O., Bøsnes, T., and Roald, L.: Regional flomfrekvensanalyse for norske vassdrag, Tech. Rep. Oslo: NVE, 1997.

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

15 Vogel, R. M. and Stedinger, J. R.: Minimum variance streamflow record augmentation procedures, Water Resources Research, 21, 715–723, 1985.