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

Thea Roksvåg[1, 2], Ingelin Steinsland[1], and Kolbjørn Engeland[3]

[1]Norwegian University of Science and Technology, NTNU, Høgskoleringen 1, 7491 Trondheim, Norway
[2]Norwegian Computing Center, NR, Postboks 114, Blindern, 0314 Oslo, Norway
[3] The Norwegian Water Resources and Energy Directorate, NVE, Middelthuns gate 29, 0368 Oslo, Norway

Correspondence: Thea Roksvåg (roksvag@nr.no)

Abstract. We present a Bayesian geostatistical model for mean annual runoff that incorporates simulations from a process-based hydrological model. The simulations are treated as a covariate and the regression coefficient is modeled as a spatial field. This way the relationship between the covariate (simulations from a hydrological model) and the response variable (observed mean annual runoff) is allowed to vary within the study area. A preprocessing step for including short records in the modeling is also suggested and we obtain a model that can exploit several data sources. By using state of the art statistical methods fast inference is achieved.

The geostatistical model is evaluated by predicting mean annual runoff for 1981-2010 for 127 catchments in Norway based on observations from 411 catchments. Simulations from the process-based HBV model on a 1 km $\times$ 1 km grid are used as input. We found that on average the proposed approach outperformed a purely process-based approach (HBV) when predicting runoff for ungauged and partially gauged catchments. The reduction in RMSE compared to the HBV model was 20 % for ungauged catchments and 58 % for partially gauged catchments, where the latter is due to the preprocessing step. For ungauged catchments the proposed framework also outperformed a purely geostatistical method with a 10 % reduction in RMSE compared to the geostatistical method. For partially gauged catchments however, purely geostatistical methods performed equally well or slightly better than the proposed combination approach. In general, we expect the proposed approach to outperform geostatistics in areas where the data availability is low to moderate.

1 Introduction

Runoff is defined as the flow of water that is generated from excess rainwater or meltwater, and that flows on the ground surface or within the soil towards a stream (WMO, 1992). Runoff indices of different types (annual runoff, seasonal runoff, maximum runoff) are needed for a variety of purposes, e.g. for designing infrastructure, water supply and hydropower reservoirs. In spite of the large importance of accurate runoff estimates, the majority of the catchments in the world are ungauged, i.e. runoff measurements for deriving the relevant indices are not available and must be predicted. This is known as the prediction of runoff in ungauged basins problem (PUB) and is a key challenge in hydrology (Blöschl et al., 2013).

There are two main approaches for predicting runoff in ungauged basins: process-based approaches and statistical approaches. In statistical approaches, data from gauged catchments are used to develop a statistical relationship between the observed runoff and relevant variables like precipitation, temperature, land use and elevation. The statistical relationship is next used to make predictions for ungauged sites with uncertainty (see e.g. Viglione et al. (2013); Merz and Blöschl (2005); Blöschl et al. (2013); Laaha and Blöschl (2005)). In this article we propose a geostatistical model for mean annual runoff (see e.g. Gelfand et al. (2010); Cressie (1993)). In the field of hydrology, several geostatistical approahces have been suggested (Roksvåg et al., 2020; Sauquet et al., 2000), but the Top-Kriging method proposed by Skøien et al. (2006) has been shown to be particularly suitable for modeling areal referenced runoff data (Viglione et al., 2013).

Process-based hydrological models are different from statistical models by that they use physical relationships for e.g. conservation of mass and energy to estimate the flow index of interest. The input variables to the process-based models are variables like precipitation, temperature and land use. Data from gauged catchments are used for validation and model calibration (see e.g. Doherty (2004); Lawrence et al. (2009)). The HBV model is an example of a process-based hydrological approach commonly used to estimate runoff in the Nordic counties (Bergström, 1976). Other process-based models are discussed in Blöschl et al. (2013); Clark et al. (2017); Fatichi et al. (2016).

The ability to account for well-known, physical relationships between the variables, is a main benefit of using a process-based model. Geostatistical approaches on the other hand provide uncertainty quantification and are typically better at ensuring a good fit between the runoff data and the model in areas where we have observations. The drawback of the geostatistical approaches are that they often depend on a relatively high gauging density and perform poorly if the underlying process is complex (Wang et al., 2017). Motivated by these benefits and drawbacks, we develop a model that combines geostatistics with a process-based approach.

There exist work based on similar ideas in the literature. In Pannecoucke et al. (2020), the authors used a process-based model to simulate flow. Next, empirical variograms were computed based on the simulations and used as input in Kriging, a class of commonly used geostatistical models (see e.g. Cressie (1993); Gelfand et al. (2010)). The goal was to estimate the contamination level within the soil . In Laaha et al. (2013), Kriging with external drift was used for interpolation of streamflow temperatures, where a physical relationship between mean annual stream temperature and stream gauge altitude was combined with the Top-Kriging approach. Qiu et al. (2018) present a model for mean annual runoff where a Budyko water balance model is combined with a geostatistical approach. In Sauquet (2006) mean annual runoff was estimated by a geostatistical method that incorporated basin characteristics through residual Kriging.

In this paper, we suggest a Bayesian model for mean annual runoff where the observed runoff is used as the response variable and where mean annual simulations from a process-based hydrological model are used as a covariate. To connect the response variable to the covariate, we use a *spatially varying coefficient* (SVC). In a model with a spatially varying coefficient, the relationship between the response variable and the covariate is allowed to vary within the study area (Gelfand et al., 2003; Ferguson et al., 2009; Hastie and Tibshirani, 1993; Su et al., 2017; Finley, 2011; Lu et al., 2009), i.e. differently from a simple linear regression model where the relationship is restricted to be constant. The motivation behind using a spatially varying

coefficient in this work, is that we assume that the process-based model is more accurate in some areas than others, and that the accuracy follows regional patterns.

There are several ways to implement a spatially varying coefficient. One option is to simply divide the study area into regions and let a given coefficient have one value for each region, as in Gamerman et al. (2003). Alternatively, the regression coefficient can be modeled as a Gaussian random field (GRF) as described in e.g. Gelfand et al. (2003). The GRF regionalizes the regression coefficient from locations with data to locations without data according to a spatial dependency structure. In this paper, we adopt the approach from Gelfand et al. (2003). In addition to the spatially varying coefficient, we include an additive spatial effect (GRF). This makes our model able to capture two different dependency structures, e.g. spatial dependency due to both short ranged and long ranged hydrological processes.

When constructing a runoff map, we find it important to exploit all available data, also data from partially gauged catchments. Partially gauged catchments are catchments that only have short records of data, from a subset of the target period. In this work, we propose to use the approach from Roksvåg et al. (2020) to preprocess the short records before further analysis with the spatially varying coefficient model. The preprocessing procedure fills in missing annual observations, and has been shown to work well for flow indices and study areas that are dominated by runoff patterns that are repeated over time Roksvåg et al. (2020). After the preprocessing, the short records are incorporated into the spatially varying coefficient model through an observation likelihood that supports data from both fully and partially gauged catchments with different observation uncertainties.

The main objective of this article is to present a framework for mean annual runoff estimation that exploits several data sources: Precipitation data, temperature data and land-use through the process-based covariate, and data from fully gauged and partially gauged catchments through the observation likelihood. The framework is made computationally feasible by using state of the art statistical methods such as INLA (integrated nested laplace approximations) and the SPDE (stochastic partial differential equation) approach to spatial modeling (Rue et al., 2009; Lindgren et al., 2011). These tools enable fast and approximate inference for Bayesian spatial models.

To evaluate the model, we estimate mean annual runoff in Norway. Simulations of mean annual runoff produced by the process-based HBV model are used as a covariate. The evaluation assess the model's ability to:

1) Produce a satisfactory gridded map for mean annual runoff with uncertainty quantification.

2) Predict runoff for partially gauged and ungauged catchments.

As reference models we use a purely process-based model (the HBV model) and a purely geostatistical model (Top-Kriging).

In the next section (Section 2), we present the available Norwegian runoff data and model input. Here, we describe the process-based HBV model and how it was used to produce simulations on a grid. In Section 3 we introduce background theory, relevant statistical models and notation. Further, in Section 4, we present the suggested mean annual runoff model. The experimental set-up for evaluating the model is presented in Section 5, and in Section 6 and 7 we present and discuss our results. Finally, we summarize and conclude in Section 8.

2 Model input

2.1 Runoff data

To evaluate the proposed approach, we use mean annual runoff data from Norway from the time period 1981-2010 provided by the Norwegian Water Resources and Energy Directorate (NVE). The mean annual runoff observations have unit mm/year and were derived by aggregating daily measurements of streamflow from Norwegian catchments, for hydrological years that starts September 1st and ends August 31st. If a catchment had less than 365 daily observations for a specific year, this annual observation was considered missing.

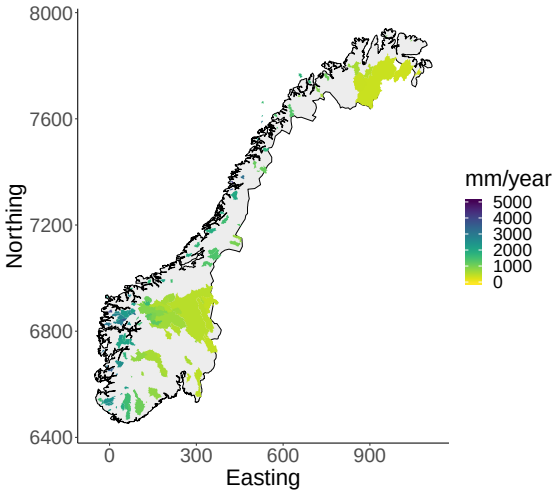
Furthermore, we only use data from catchments where human activities have had negligible impact. To select catchments, we used the regulation capacity of hydropower reservoirs as a criterion, i.e. the ratio between the mean annual runoff and the reservoir storage capacity. If this ratio was smaller than 0.2 for a catchment, the catchment was included in the analysis, assuming that the annual changes in water storage is small compared to the annual inflow volume. The assumption was also checked for a subset of the target catchments, and we found that the standard deviation of annual changes in reservoir storage was less than 2 % of annual inflow.

After performing the data cleaning procedure, there were data available from 127 catchments that were *fully gauged* in the 30 year target period, 1981-2010. The average runoff for these catchment are shown in Figure 1a with unit mm/year. In addition, there were annual observations available from 284 *partially gauged catchments*. These had at least 1 annual runoff observation between 1965 and 2010 and their observed mean annual runoff is shown in Figure 1b. The number of annual observations for each of these catchments is shown in Figure 1c. The average record length is 12 years (median 9.5 years) for 1981-2010, but 15 years (median 16 years) if we consider the longer time period from 1965 to 2010.

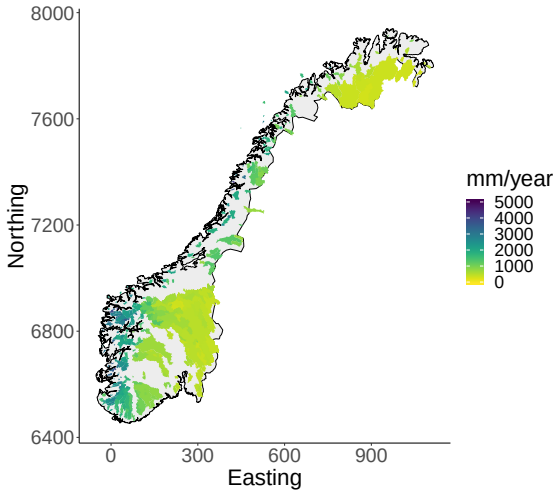
In Figure 1d we show the annual runoff observations for individual years. Here, it is apparent that the spatial variability of the Norwegian annual runoff is large. The mean annual runoff follows the spatial pattern we see in Figure 1b, with large observations in the western part of the country and smaller observations in eastern part each year. The pattern is mainly caused by the orographic precipitation that occurs when humid winds from the Atlantic ocean are elevated over the mountains in western Norway. This gives large precipitation amounts in the western parts of the country, while the eastern parts are left in the rain shadow.

2.2 Gridded simulations from the HBV model

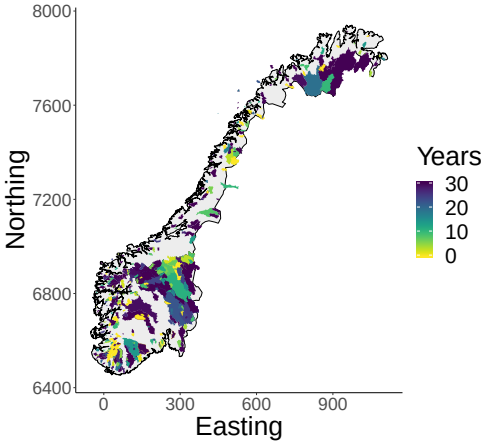
In this study, we use a gridded mean annual runoff product simulated by the HBV model as a covariate in a geostatistical model. The first application of the gridded HBV model in Norway is reported in Beldring et al. (2002), and it is applied in several studies to assess runoff and water balance in Norway (e.g. in Borgvang et al. (2006); Skarbøvik et al. (2009); Hanssen-Bauer et al. (2017)). In this case, we use a data product that was already available from the data provider NVE's database. See Figure 2a. The product was delivered on a 1 km $\times$ 1 km grid and is based on simulations of daily time series of runoff. Interpolated temperature and precipitation were used as input together with gridded land use characteristics. Daily simulated time series of runoff were aggregated to mean annual runoff (mm/year) for our reference period 1981-2010. We refer Bergström (1976);



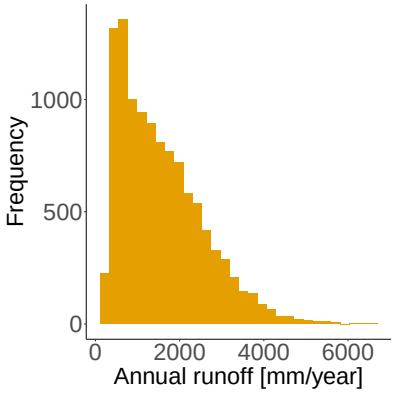
(a) Catchmens that are fully gauged in the study period (1981-2010).



(b) Fully and partially gauged catchments (1965-2010).

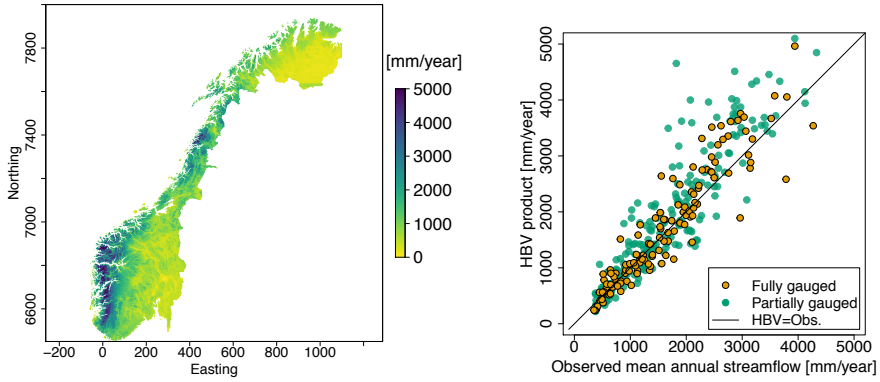


(c) Number of annual observations available for 1981-2010.



(d) Observed annual runoff from fully gauged and partially gauged catchments (1965-2010).

Figure 1. Mean annual runoff for Norwegian catchments (upper plots) derived from daily streamflow observations. There are annual runoff data available from 127 fully gauged catchments and from 284 partially gauged catchments. We plot subcatchments in front of larger, surrounding catchments in all of our plots. Figure 1c shows the number of annual observations that are available for each catchment in the study period (1981-2010). If the number is equal to 0, it means that there is at least one annual observation available from 1980 or earlier years, more specifically between 1965 and 1980. The reference system used is UTM33N EUREF89 with coordinates given in km. Figure 1d shows the annual observations available for all catchments and years.



(a) Gridded product from the HBV model.

(b) HBV versus observed streamflow.

Figure 2. A mean annual runoff (1981-2010) product simulated by the HBV model (Figure 2a). The product is delivered on a 1 km $\times$ 1 km grid. Figure 2b shows the fit between the HBV product and the actually observed streamflow for the fully gauged (orange) and partially gauged catchments (green).

Sælthun (1996); Lindström et al. (1997) detailed descriptions of the algorithms used in the HBV model, and to Beldring et al. (2002) for details about the specific product in Figure 2a.

To determine the parameters in the HBV model, it is common to perform a global calibration procedure based on streamflow observations. The aim is to minimize global bias and the errors. When producing the map in Figure 2a, streamflow observations from 141 fully gauged catchments were used for the calibration. Remark that since we use the HBV product that already was available from the data provider NVE's database, the calibration catchments are not necessarily the same catchments we use in our geostatistical model. See Beldring et al. (2003) for details about the calibration procedure.

As the parameters are calibrated globally, there are still local biases in the HBV model's predictions relative to the observed streamflow. This can be seen in Figure 2b where we visualize the difference between the mean annual runoff provided by the HBV model (Figure 2a) and the observed mean annual runoff (Figure 1). We see that the fit is close to linear for catchments with low observations of mean annual runoff. For observations over 2000 mm/year, the HBV model tends to overestimate the mean annual runoff. By using the proposed geostatistical approach, we aim to produce a runoff map that improves the fit.

3 Methodological background

3.1 Bayesian statistics and hierarchal modeling

We take a Bayesian approach to statistics (see e.g.Gelman et al. (2004); Casella and Berger (1990)). In Bayesian statistics, the random variable $\boldsymbol{x}$ is associated with a probability distribution that expresses what we know about the underlying process of interest. Before the statistical analysis is conducted, our beliefs are expressed mathematically through a so-called prior distribution, denoted $\pi(\boldsymbol{x})$. This can be constructed based on expert knowledge about the process under study or based on

earlier experiments. The goal of the Bayesian analysis, is to update $\pi(\boldsymbol{x})$ based on data $\boldsymbol{y}$. This can be done by using Bayes' formula:

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

where $\pi(\boldsymbol{y}|\boldsymbol{x})$ is the observation likelihood that connects the observed values $\boldsymbol{y} = (y_1, ..., y_m)$ to the target variable $\boldsymbol{x}$. The resulting distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ is called the posterior distribution, and represents what we know about the underlying process based on our data. One of the benefits the Bayesian framework, is that a full uncertainty specification for the target variable $\boldsymbol{x}$ is directly available through the posterior distribution. If a point prediction is of interest, the median, mean or mode of the posterior distribution $\pi(x_i|\boldsymbol{y})$ can be used as a summary statistic, for any $x_i \in \boldsymbol{x}$.

The geostatistical runoff model we propose, is also a *hierarchical* model. Hierarchical models make it possible to formulate rather complex models by specifying a set of simpler models (see e.g. (Banerjee et al., 2004)). For example if we model runoff, we can assume that the true underlying runoff $\boldsymbol{x}$ is observed through data $\boldsymbol{y}$ that are associated with some measurement uncertainty. Further can we assume that the runoff has some spatial or temporal variability that can be modeled by a statistical distribution with parameters $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)$. Mathematically can the above model be expressed in three stages: The *observation likelihood*, $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$, the *latent model* or *process model* $\pi(\boldsymbol{x}|\boldsymbol{\theta})$ and the prior distributions $\pi(\boldsymbol{\theta})$.

3.2 Gaussian random fields (GRFs)

Random fields (RFs) are often used to model spatial correlation in geostatistical models for hydrological variables (see e.g. Sauquet et al. (2000); Skøien et al. (2006); Roksvåg et al. (2020)). In this article, we use Gaussian random fields (GRFs) to model runoff. A continuous field $\{x(\boldsymbol{u}); \boldsymbol{u} \in \mathcal{D}\}$ defined on a spatial domain $\mathcal{D}$ is a Gaussian random field if $(x(\boldsymbol{u}_1), ..., x(\boldsymbol{u}_n))^{\mathrm{T}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\mathcal{N}(\cdot, \cdot)$ is a multivariate normal distribution with expected values given by vector $\boldsymbol{\mu}$ and covariance given by the covariance matrix $\boldsymbol{\Sigma}$ (Cressie, 1993). The covariance matrix specifies the dependency structure of the variable of interest. Typically, a matrix element (i, j) is generated by using a known covariance function $C(\boldsymbol{u}_i, \boldsymbol{u}_j)$ that models the correlation of the target variable between two locations $\mathrm{Cov}(x(\boldsymbol{u}_i), x(\boldsymbol{u}_j))$. The covariance function typically has a marginal variance parameter σ^2 and a range parameter ρ that characterize the underlying spatial field. The marginal variance describes the spatial variability of the target variable, while the range is a measure of how correlation decays with distance.

In our work we use a stationary Matérn covariance function to model the covariance of mean annual runoff. The Matérn covariance function is defined as:

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

where K_ν is the modified Bessel function of second kind and order $\nu > 0$, $\Gamma(\cdot)$ is the gamma function and $||\boldsymbol{u}_j - \boldsymbol{u}_i||$ is the Euclidean distance between the two locations $\boldsymbol{u}_i, \boldsymbol{u}_j \in \mathcal{R}^d$. Further, is σ^2 the marginal variance and κ is a scale parameter (Guttorp and Gneiting, 2006). Empirically, it has been shown that the parameters ν and κ can be used to express the spatial range as

$$\rho = \sqrt{8\nu}/\kappa, \tag{3}$$

where ρ is defined as the distance at which the correlation between two locations has dropped to 0.1 (Rue et al., 2009).

The reason for using a Matérn covariance function in our work, is that it makes it possible to use the stochastic partial differential equation (SPDE) approach to spatial modeling (Lindgren et al., 2011). The SPDE approach is described in Section 4.6 and is used to make the proposed model computationally feasible. In addition, the Matérn class of covariance functions has many useful properties and Stein (1999) advice to use it.

3.3 Existing geostatistical models used for runoff interpolation

3.3.1 Top-Kriging

Kriging approaches are commonly used for spatial interpolation. In Kriging approaches, the variable of interest is modeled as a random field $x(\boldsymbol{u})$, and an estimate of the random field $x(\boldsymbol{u}_0)$ at an unobserved location $\boldsymbol{u}_0 \in \mathcal{R}^2$ can be expressed as the weighted sum of a set of observations $x(\boldsymbol{u}_i), ..., x(\boldsymbol{u}_n)$, i.e. as

$$\hat{x}(\boldsymbol{u}_0) = \sum_{i=1}^{n} \lambda_i x(\boldsymbol{u}_i), \tag{4}$$

where λ_i for $i = 1, ..n$ are interpolation weights that must be determined (Cressie, 1993). The interpolation weights are found by minimizing the mean squared error between the estimate $\hat{x}(\boldsymbol{u}_0)$ and the true $x(\boldsymbol{u}_0)$, and assuming zero mean expected error. A linear estimator with these properties is called the best linear unbiased estimator (BLUE).

The estimation of the Kriging weights requires evaluations of the covariance function (or variogram) of the involved random field, and the covariance typically depends on the distance between the observation locations $\boldsymbol{u}_i$. However, runoff observations are linked to catchment areas rather than to single point locations. In the Top-Kriging approach (Skøien et al., 2006), this is taken into account by treating the runoff observations as areal referenced when computing the covariance. This makes it possible to weight an observation from a subcatchment more than observations from a nearby, non-overlapping catchments. Top-Kriging is one of the leading methods for interpolation of runoff (Viglione et al., 2013; Blöschl et al., 2013), and we hence use it as a geostatistical reference method.

3.3.2 Geostatistical method for exploiting short records

To include short records in our model, we use the method from Roksvåg et al. (2020) as a preprocessing step. The method is a Bayesian hierarchical geostatistical model that is particularly suitable for filling in missing data for catchments that have short records of data relative to neighboring catchments. It models several years of (annual) runoff simultaneously through two GRFs: one that describes the long-term spatial variability, and one that describes year dependent spatial effects. The method weights the two GRFs relative to each other. If long-term effects dominate, the potential information stored in short records is large.

The method from Roksvåg et al. (2020) has its benefits when runoff follows spatial patterns that are repeated over time. This is the case for our target variable, Norwegian annual runoff, that is driven by orographic precipitation caused by repeated wind patterns from the Atlantic ocean (Stohl et al., 2008). Example data from Norway are shown in 3. The repeated spatial pattern

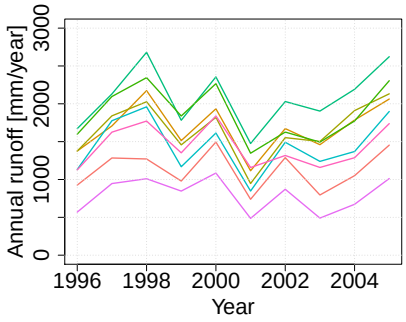


Figure 3. Time series of annual runoff for 8 catchments in Norway that are located in the same region. The time series are almost parallel, indicating that the spatial patterns of runoff are repeated over time.

is recognized by that the ranking of the catchments, from wet to dry, is approximately constant. For variables and areas that are not driven by such characteristic spatial patterns, the method in Roksvåg et al. (2020) provides a more classical form of spatial interpolation, similar to Kriging.

The method from Roksvåg et al. (2020) is available for both point and areal referenced data. In this application, we use it as a point referenced model to save computational time, and the catchments centroids are used as the observation locations. We expect the point referenced model to be sufficiently good for our study for two reasons: 1) We are only using the model to make predictions for catchments where we have at least one annual observation and 2) we are not going to use the posterior uncertainty of the model. The results in Roksvåg et al. (2020) show that the point referenced model gives results that are similar to the areal referenced model for partially gauged catchments when we only are interested in posterior means and not posterior standard deviations.

4 A spatially varying coefficient (SVC) model for incorporating process-based simulations and short records

We now present the proposed geostatistical Bayesian hierarchical model for mean annual runoff and its three stages: A process model, an observation model and prior distributions.

4.1 Process model for true mean annual runoff

4.1.1 Point model

Assume that mean annual runoff (mm/year) is a continuous process that occurs for any point $\boldsymbol{u} \in \mathcal{R}^2$ in the landscape. We model the true mean annual runoff $q(\boldsymbol{u})$ at a point location or a (small) grid cell $\boldsymbol{u}$ as

$$
\begin{aligned}
&q(\boldsymbol{u}) = \beta_0 + (\beta_1 + \alpha(\boldsymbol{u})) \cdot h(\boldsymbol{u}) + x(\boldsymbol{u}); && \text{(SVC model)} \\
&x(\boldsymbol{u})|(\rho_x, \sigma_x) \sim \mathrm{GRF}(\rho_x, \sigma_x) && \beta_0 \sim \mathcal{N}(0, (10000 \text{ mm/year})^2) \\
&\alpha(\boldsymbol{u})|(\rho_\alpha, \sigma_\alpha) \sim \mathrm{GRF}(\rho_\alpha, \sigma_\alpha) && \beta_1 \sim \mathcal{N}(0, 10000^2)
\end{aligned}
\tag{5}
$$

where β_0 is an intercept with a normal distributed prior. The variable $h(\boldsymbol{u})$ is a covariate that contains the simulated value generated by a process-based hydrological model at point location or a grid cell $\boldsymbol{u}$, and $(\beta_1 + \alpha(\boldsymbol{u}))$ defines a spatially varying coefficient (SVC). The spatially varying coefficient consists of one fixed effect β_1 and one component $\alpha(\boldsymbol{u})$ that changes in space. The spatial variability of $\alpha(\boldsymbol{u})$ is introduced by modeling it as a stationary Matérn Gaussian random field given a range parameter ρ_α and a marginal standard deviation parameter σ_α. This way the relationship between the true mean annual runoff $q(\boldsymbol{u})$ and the simulations made by the hydrological model $h(\boldsymbol{u})$ can vary in the study area. The $\alpha(\boldsymbol{u})h(\boldsymbol{u})$ component also ensures a model where the mean and the variance of runoff can be inhomogeneous in space. For the fixed effect β_1 we use a weakly informative normal prior distribution with zero mean and standard deviation 10000. This is the same prior as for the intercept β_0. We use mean annual runoff simulations from the HBV model as input in $h(\boldsymbol{u})$, but gridded simulations from any relevant hydrological model can be applied.

The spatially varying coefficient $\beta_1 + \alpha(\boldsymbol{u})$ in Equation (5) models a similar dependency structure as we would get from *ratio interpolation*, i.e. interpolation of the ratio between the observed runoff and a process-based covariate. Ratio interpolation is a method that has been used before in e.g. Beldring et al. (2002) to improve the results of a process-based model. In our runoff model, we also include an additional spatial effect $x(\boldsymbol{u})$ that is assumed to be conditionally independent of $\alpha(\boldsymbol{u})$. Like $\alpha(\boldsymbol{u})$, $x(\boldsymbol{u})$ is modeled as a GRF with a stationary Matérn covariance structure, but with range and marginal standard deviation ρ_x and σ_x respectively. The GRF $x(\boldsymbol{u})$ models a dependency structure similar to what we would get from *residual interpolation*. Residual interpolation was used in e.g. Merz and Blöschl (2005) to improve the results from an initial multiple linear regression model.

4.1.2 Areal model

In Equation (5) we modeled runoff as a point referenced process, but in practice, runoff is observed through streamflow observations that are linked to catchment areas. We thus introduce a model for the true runoff inside a catchment area $\mathcal{A}$. This is given by:

$$
Q(\mathcal{A}) = \frac{1}{|\mathcal{A}|} \int_{\boldsymbol{u} \in \mathcal{A}} q(\boldsymbol{u}) d\boldsymbol{u}, \tag{6}
$$

where $q(\boldsymbol{u})$ is the mean annual point runoff from Equation (5) and $|\mathcal{A}|$ is the area of the target catchment. The true areal runoff is given by the average point runoff integrated over the catchment area. In practice, it is not computationally feasible to perform the integration in Equation (6). Our solution is to approximate the integral in Equation (6) by a sum. This is done by discretizing catchment $\mathcal{A}$ into a regular grid $\mathcal{L}_{\mathcal{A}}$ and defining the mean annual runoff in catchment $\mathcal{A}$ as:

$$Q(\mathcal{A}) \approx \frac{1}{n_{\mathcal{A}}} \sum_{\boldsymbol{u} \in \mathcal{L}_{\mathcal{A}}} q(\boldsymbol{u}), \tag{7}$$

where $n_{\mathcal{A}}$ is the number of grid nodes in the discretization of catchment $\mathcal{A}$. The areal formulation in Equation (7) assumes a linear aggregation of runoff over the grid nodes included in the catchment discretization. This is reasonable for variables that are approximately mass conservative, like the mean annual runoff.

We have now defined our final process model for runoff, which is an areal model (Equation (7)) that builds on a point specification of the underlying process (Equation (5)). From equations (5) and (7), we see that in order to calculate $Q(\mathcal{A})$ we have to evaluate $h(\mathcal{A}) = \sum_{\boldsymbol{u} \in \mathcal{L}_{\mathcal{A}}} h(\boldsymbol{u})$, i.e. we need the simulated values produced by the hydrological product $h(\boldsymbol{u})$ for the grid nodes inside $\mathcal{A}$. Consequently, the catchment discretization should follow the same discretization as the gridded hydrological product that is used as input to $h(\boldsymbol{u})$. In our case the HBV product comes on a regular grid with 1 km spacing. The selected grid should be dense enough to ensure an accurate approximation for the true areal runoff in Equation (7).

4.2 Observation model for mean annual runoff

The true mean annual $Q(\mathcal{A})$ runoff is not observed directly, but through areal referenced streamflow observations with uncertainty. We model the observed mean annual runoff in catchment $\mathcal{A}_i$ as:

$$y_i = Q(\mathcal{A}_i) + \epsilon_i, \tag{8}$$

where $Q(\mathcal{A}_i)$ is the areal referenced true mean annual runoff from Equation (7), and the ϵ_i's are independent and identically distributed error terms with prior $\mathcal{N}(0, s_i\sigma_y^2)$. The parameter σ_y describes the underlying standard deviation, while the s_i's are fixed, predetermined scales that allow each observation to have its own measurement uncertainty. This way heteroscedasticity can be introduced in a simple way. The values of the scales s_i are further specified in Section 4.3.

In Equation (8), all components are Gaussian, which means that there is a risk of obtaining negative runoff predictions from the proposed model. To avoid negative runoff predictions, we could log transform the runoff data before performing the analysis, but this requires that we model the runoff observations as point referenced instead of areal referenced. The reason is that the sum in Equation (7) does not make sense for log transformed runoff data. Another option is to use a log-Gaussian likelihood and log-Gaussian random fields for $x(\boldsymbol{u})$ and $\alpha(\boldsymbol{u})$, such that predictions for $x(\boldsymbol{u})$ and $\alpha(\boldsymbol{u})$ always are positive. In this work, however, we keep the areal formulation and the more interpretive versions of the spatial fields. For Norwegian mean annual runoff, negative predictions are quite unlikely anyway, since the observations are far away from zero. The areal formulation also gives a more realistic uncertainty model and let us constrain the mean annual runoff not only at certain gauging points, but over catchment areas.

4.3 Prior distributions for model parameters

The third stage of the proposed hierarchical model for mean annual runoff consists of the prior distributions of the 5 model parameters, $(\rho_\alpha, \sigma_\alpha, \rho_x, \sigma_x, \sigma_y)$. In this section we specify our prior distributions. Most of the priors are constructed such that they are suitable for modeling Norwegian mean annual runoff, and should be revised the model is used for other flow indices and/or study areas.

We start by constructing a prior for the measurement uncertainty $s_i\sigma_y^2$. As stated in the previous subsection, the variance parameter σ_y^2 is scaled with a fixed an predetermined scale s_i such that each observation of mean annual runoff can have its own measurement uncertainty. A variance that changes with the observed value is reasonable when modeling Norwegian mean annual runoff, because the variability of runoff across the country is large. The observed annual runoff varies from around 500 mm/year to 4000 mm/year. With this in mind, we specify the scales s_i under the assumption that larger observations of mean annual runoff have larger measurement uncertainties than smaller observations of mean annual runoff. This is obtained by modeling the scales as

$$s_i = (0.025 \cdot y_i/1000)^2, \tag{9}$$

where y_i is the observed mean annual runoff in catchment $\mathcal{A}_i$ in mm/year. The number 0.025 was chosen according to expert opinions from the data provider NVE. A standard deviation around 2.5 $\%$ is assumed to be reasonable. The scales are divided by a factor of 1000 to get suitable values for the quantity $s_i \cdot \sigma_y^2$.

We next specify a prior distribution for the standard deviation parameter σ_y. For this, we use a penalized complexity (PC) prior as suggested by Simpson et al. (2017). The PC prior is chosen because it has convenient mathematical properties. It controls for overfitting by penalizing the increased complexity that arises when a model deviates from a simpler, less flexible base model. The PC prior for the precision τ (or the inverse variance) of a Gaussian effect $\mathcal{N}(0, \tau^{-1})$ is given by

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

where λ controls the deviation penalty. The parameter λ can easily be specified through a probability α and a quantile u as $\text{Prob}(\sigma > \sigma_0) = \alpha$, where $\sigma_0 > 0$, $0 < \alpha < 1$ and $\lambda = -\ln(\alpha)/u$, where $\sigma = 1/\sqrt{\tau}$ is the standard deviation of the Gaussian effect. For our application, we let $\alpha = 0.1$ and $\sigma_0 = 1500$ mm /year, and define the PC prior for σ_y as follows:

$$\text{Prob}(\sigma_y > 1500 \ \ \text{mm}) = 0.1. \tag{11}$$

This means that the prior probability that σ_y is larger than 1500 mm/year is 10 $\%$. However, recall that the measurement variance of y_i is determined by $s_i\sigma_y^2$ and not by σ_y^2 alone. With the scales in Equation (9) and the PC prior for σ_y in Equation (11), a prior 95% credibility interval for the observation standard deviation for the mean annual runoff is $(0.04, 6)\%$ of the corresponding observed value y_i for a catchment $\mathcal{A}_i$, with prior mean centered around 2.5%. Values in this range are reasonable and reflect the data provider NVE assumptions about the uncertainty of the Norwegian mean annual runoff observations. By creating a relatively narrow prior for $s_i\sigma_y^2$, we influence the model to reproduce the actually observed runoff for catchments where we have data.

In Fuglstad et al. (2019) the PC prior framework is used to develop a informative, joint prior for the range and the marginal variance of a Gaussian random field. We use this prior for the spatial marginal standard deviation σ_α and the spatial range ρ_α for the spatially varying coefficient component $\alpha(\boldsymbol{u})$. The prior is specified through the following probabilities and quantiles

$$\text{Prob}(\rho_\alpha < 20 \text{ km}) = 0.1, \qquad \text{Prob}(\sigma_\alpha > 2) = 0.1, \tag{12}$$

where we a priori assume that the spatial range of the spatially varying coefficient is larger than 20 km. This is a reasonable assumption, as it is likely that locations that are closer than 20 km are correlated when it comes to annual runoff. Based on Figure 2b we assume a prior that the ratio between the response variable $Q(\cdot)$ and the covariate $h(\cdot)$ varies with a factor that has a standard deviation smaller than 2.

Likewise, we use the PC prior from Fuglstad et al. (2019) to specify a joint prior for the marginal standard deviation σ_x and the spatial range ρ_x of the spatial field $x(\boldsymbol{u})$. We use the following probabilities and quantiles:

$$\text{Prob}(\rho_x < 20 \text{ km}) = 0.1, \qquad \text{Prob}(\sigma_x > 2000 \text{ mm/year}) = 0.1. \tag{13}$$

Here, we again assume a prior that the range is larger than 20 km by taking the size of the study area into account. The prior probability that the standard deviation of the Norwegian mean annual runoff is larger than 2000 mm/year is set to a low probability. We find this reasonable as most of the mean annual observations are between 500 mm/year and 4000 mm/year.

4.4 Preprocessing step for incorporating short records (PP)

We now present an extension of the model that makes it possible to include short records in the proposed mean annual runoff model. The extension is based on using the geostatistical model described in Section 3.3.2 to fill in missing annual observations and/or augment short records for the partially gauged catchments. After filling in the missing years, we get a preliminary estimate of the mean annual runoff for these catchments. These estimates are next used as observations y_i in the SVC model together with data from fully gauged catchments.

The observations y_i we obtain from the preprocessing step are probably more uncertain than the data from the fully gauged catchments. To reflect this, we use a different prior for the observation uncertainty for the preprocessed data compared to that of the fully gauged catchments'. Recall that the prior observation variance for a fully gauged catchment was given by $s_i\sigma_y^2$ where s_i was a fixed predetermined scale given by $s_i = (0.025 \cdot y_i/1000)$. For partially gauged catchments we replace this scale by

$$s_i^{\text{PP}} = (0.10 \cdot y_i/1000), \tag{14}$$

where PP denotes that observation y_i from catchment A_i is preprocessed. In practice, each partially gauged catchment could have its own scaling factor, but in this demonstration we use the same scaling factor for all partially gauged catchments for simplicity. With the scales in Equation (14), a 95 % credible interval for the prior standard deviation $\sqrt{s_i\sigma_y^2}$ becomes $(0.1, 24)$ % of the observed value for the partially gauged catchments, while it is only (0.04,6) % for data from fully gauged catchments.

The preprocessing step let us exploit streamflow observations from catchments that have down to one annual observation, and the short record could also be from the period before the study period starts. As explained in Section 3.3.2 the preprocessing step is expect to contribute positively to the model if the flow index of interest is driven by repeated spatial patterns over time. If this is not the case, the preprocessing step only performs classical geostatistical spatial interpolation and can be skipped to save time.

4.5 Full model specification

We have proposed a model for mean annual runoff that can incorporate process-based simulations and data from fully gauged and partially gauged catchments. The full model can be specified in as a hierarchical model with three levels, where the first level is he observation likelihood,

$$\pi(\boldsymbol{y}|\boldsymbol{x},\sigma_y) = \prod_{i=1}^{n}(I\{\text{Catchment } \mathcal{A}_i \text{ is fully gauged}\} \cdot \mathcal{N}(Q(\mathcal{A}_i), s_i\sigma_y^2) + I\{\text{Catchment } \mathcal{A}_i \text{ is partially gauged}\} \cdot \mathcal{N}(Q(\mathcal{A}_i), s_i^{\text{PP}}\sigma_y^2), \tag{15}$$

the second level is the latent field,

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}) = \pi\left(x(\boldsymbol{u_1}),...,x(\boldsymbol{u_m})|\rho_x,\sigma_x\right) \cdot \pi\left(\alpha(\boldsymbol{u_1}),...,\alpha(\boldsymbol{u_m})|\rho_\alpha,\sigma_\alpha\right) \cdot \pi(\beta_0) \cdot \pi(\beta_1), \tag{16}$$

and the third level is the prior model,

$$\pi(\sigma_y,\boldsymbol{\theta}) = \pi(\rho_x,\sigma_x) \cdot \pi(\rho_\alpha,\sigma_\alpha) \cdot \pi(\sigma_y). \tag{17}$$

Above, $\boldsymbol{y}$ is a vector containing all observations $y_i,...,y_n$ of mean annual runoff for catchments $\mathcal{A}_1,...,\mathcal{A}_n$. The function $I(\cdot)$ is an indicator function that is equal to one if its argument is true and zero otherwise, allowing for data from both fully gauged and partially gauged catchments. We see that the likelihood specification for the fully and partially gauged catchments is the same, except for the difference in measurement uncertainty expressed through s_i and s_i^{PP}. The variable $\boldsymbol{x}$ is a vector that contains all the latent variables, i.e. the two fixed effects β_0 and β_1, and the two Gaussian random fields $x(\boldsymbol{u}_1),...,x(\boldsymbol{u}_m)$ and $\alpha(\boldsymbol{u}_1),...,\alpha(\boldsymbol{u}_m)$ for all grid nodes $\boldsymbol{u}_1,..,\boldsymbol{u}_m$ that are used in the discretization of the catchment areas. Finally is $\boldsymbol{\theta}$ a parameter vector that contains $\rho_x,\sigma_x,\rho_\alpha$ and σ_α. Together with σ_y it defines all the model parameters.

In Figure 4 we visualize the proposed approach in a flow chart. We emphasize that the SVC model can be used with or without incorporating preprocessed short records. To mark results where preprocessed data are involved, we will use the subscript PP in the remainder of the paper.

4.6 Approximate inference

To make runoff predictions $q(\boldsymbol{u})$ and $Q(\mathcal{A})$, we need to estimate $\boldsymbol{x}$ and $\boldsymbol{\theta}$ given data $\boldsymbol{y}$. Traditionally, inference on Bayesian hierarchical models has been done by using Markov Chain Monte Carlo (MCMC) methods (Gamerman and Lopes, 2006).

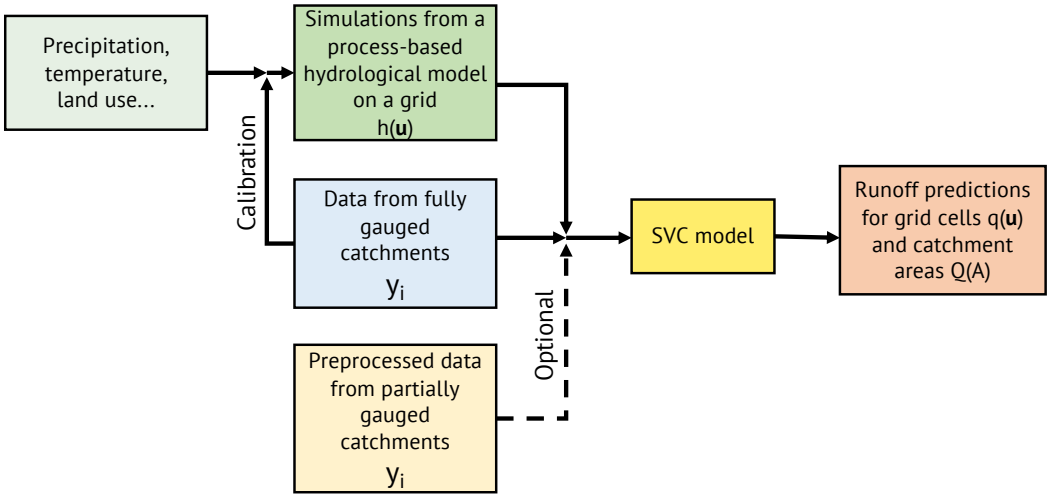


Figure 4. Workflow for estimating runoff for grid nodes $q(\boldsymbol{u})$ and for catchment areas $Q(\mathcal{A})$.

However, the computational complexity of carrying out a MCMC procedure is large when the dimension of $\boldsymbol{x}$ is large. To make the proposed model computationally feasible, integrated nested Laplace approximations (INLA) are used. The INLA methodology was suggested by Rue et al. (2009) and can be used for making approximate Bayesian inference on latent Gaussian models (LGMs), i.e. hierarchical models where the latent field $\boldsymbol{x}$ is Gaussian. As the latent variables contained in $\boldsymbol{x}$ are given Gaussian prior distributions given the model parameters, this requirement is fulfilled for our SVC model. The INLA methodology is based on Laplace approximations, sparse matrix calculations and numerical integration schemes, and we refer to Rue et al. (2009) for details.

Furthermore, it is computationally challenging to make statistical inference on spatial models. The reason is that it takes time to do matrix operations on the covariance matrices of GRFs when there are many target locations. To ensure fast inference for our two field model, we use the SPDE approach to spatial modeling, as suggested by Lindgren et al. (2011). The approach is based on the fact that a GRF with Matérn covariance matrix can be expressed as the solution of a stochastic partial differential equation (Whittle, 1954, 1963). An approximate solution of the SPDE can be obtained by using the finite element method (see e.g. Brenner and Scott (2008)), where the resulting approximation is given on a triangular mesh. This mesh approximation gives computational benefits compared to the exact GRF solution, and enables fast inference for spatial models (Rue and Held, 2005; Rue et al., 2009).

The INLA and SPDE methodologies are implemented in the r-package `INLA`, which since its introduction has been used within a range of different fields. See Opitz et al. (2018); Guillot et al. (2014); Myrvoll-Nilsen et al. (2020); Bakka et al. (2018); Blangiardo and Cameletti (2015); Khan and Warner (2018) and `www.r-inla.org` for some examples. The approximations used in the SPDE and INLA framework are in general accurate and reliable when the likelihood is Gaussian, as in this ap-

plication, and as long as the triangular mesh used in the finite element computations is dense enough relative to the spatial variability of the target variable. A mesh that is too coarse can lead to unrealistic results such as negative runoff.

5 Experimental set-up and evaluation scores

5.1 Making a gridded mean annual runoff map for 1981-2010

To evaluate the proposed approach for runoff estimation, we use the SVC model to produce a gridded mean annual runoff map for 1981-2010 for the same 1 km $\times$ 1 km grid as the HBV model was delivered on (Figure 2a). For the fully gauged catchments, we use the data from 1981-2010 to compute the mean annual runoff y_i, while for the partially gauged catchments we use the preprocessing step on the short records (PP) before further analysis. In the preprocessing step, data from 1965-2010 were used to estimate the mean annual runoff for 1981-2010.

We evaluate the model in terms of whether the new map represents an improvement compared to the original HBV map. This is done by investigating how well the new map fits with the observed runoff from the fully gauged and partially gauged catchments.

In addition to the experiment described above, we repeated the experiment, but omitted partially gauged catchments and short records from the analysis. This was done to show that the SVC model works regardless of the preprocessing step. These results can be found in Appendix A.

5.2 Cross-validation for ungauged and partially gauged catchments

We next evaluate the framework's ability to perform accurate mean annual runoff for ungauged and partially gauged catchments. This is done by a cross-validation assessment, where we do predictions for the 127 fully gauged catchments in Figure 1a. The 127 fully gauged catchments are divided into five groups or folds. The four first folds have 25 so-called target catchments, while the fifth fold has 27 target catchments. In turn, the streamflow data corresponding to each fold are removed from the dataset, while the remaining observations are used to predict the mean annual runoff for these catchments for 1981-2010. The likelihood consists of preprocessed observations from partially gauged catchments and observations from fully gauged catchments, i.e. around 400 observation catchments in total. Remark that we don't calibrate the HBV model for each cross-validation fold, as the HBV product was a pre-made product.

In our evaluation, we compare the predictive performance of the SVC model with the process-based HBV model. The original simulations from the HBV model shown in Figure 2a are hence used as they are. For evaluation purposes, the values in Figure 2a are aggregated and averaged to catchment runoff for the catchments in Figure 1.

We also compare our approach to the purely geostatistical Top-Kriging (TK) approach described in Section 3.3. Here, we fit a covariance model based on a multiplication of a modified exponential and fractal variogram model to the mean annual runoff data. This is the default variogram model in the R package `rtop` (Skøien, J.O., 2018). As for the SVC model, data from both fully gauged catchments and preprocessed partially gauged catchments are used as input, and we mark the Top-Kriging

results by TK_{PP} to emphasize that preprocessed data are used. For fully gauged catchments, the standard deviation of the observations is set to 2.5 % of the observed value y_i in the Top-Kriging approach, while for partially gauged catchments the standard deviation is set to 10 % of the observed value. The aim is to make the Top-Kriging results as comparable as possible to the SVC model results.

In addition to evaluating Top-Kriging and the HBV model, we include prediction results from the preprocessing step (PP) alone, without performing any further analysis. The PP predictions come from the purely geostatistical method described in Section 3.3.2. We include the PP results to make the Top-Kriging and SVC results more transparent. These methods use the PP results as input data for the partially gauged catchments (see Section 4.4).

The described cross-validation procedure is first performed when the 127 target catchments are treated as ungauged. We have the following setting:

Ungauged catchments (UG): The target catchments in each cross-validation fold are treated as totally ungauged (UG) in the time period of interest (1981-2010) and their observations are removed from the dataset. Observations from fully gauged catchments from other cross-validation groups and observations from partially gauged catchments are used to make predictions.

We also evaluate the predictive performance of the model when the 127 target catchments are treated as partially gauged by doing the following experiment:

Partially gauged catchments (PG): The target catchments in each cross-validation are allowed to have 3 annual observations in the study period (1981-2010). These are randomly drawn from years 1981-2010. The remaining 27 observations are removed (and observations from before 1981). Observations from nearby fully gauged catchments (1981-2010) and partially gauged neighboring catchments (1965-2010) are included in the likelihood as before.

The same cross-validation groups are used for all experiments, such that the results become comparable across methods. The randomly drawn short records of length 3 are also the same for Top-Kriging and the SVC approach.

In addition to the above experiments, we carried out a cross-validation for the UG setting when omitting catchments with short records and preprocessed data. These results can be found in Appendix A.

5.3 Evaluation scores

To evaluate the accuracy of the predictions obtained from the cross-validation, we use three evaluation scores. These are the root mean square error (RMSE), the absolute normalized error (ANE) and the Nash-Sutcliffe model efficiency coefficient (NSE), which are defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{Q}(\mathcal{A}_i))^2}, \tag{18}$$

$$\mathrm{ANE}_i = \frac{|y_i - \hat{Q}(\mathcal{A}_i)|}{y_i}, \tag{19}$$

and

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{n}(\hat{Q}(\mathcal{A}_i) - y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}. \tag{20}$$

Here, $\hat{Q}(\mathcal{A}_i)$ is the predicted mean annual runoff in catchment $\mathcal{A}_i$, y_i is the corresponding observed value and $\overline{y}$ denotes the average observed mean annual runoff over all study catchments $i = 1, ...n$. For the suggested SVC model, we use the posterior mean of $Q(\mathcal{A})$ as the predicted value (Equation (7)). As a summary statistic for ANE_i, we use the average ANE_i over all catchments $i = 1, ..n$. A low average ANE_i or a low RMSE corresponds to accurate predictions. The NSE on the other hand takes values between $-\infty$ and 1, and the closer the model efficiency is to 1, the more accurate the model is. The ANE and the NSE are different from the RMSE in being scale-independent evaluation scores.

The three above scores are suitable for evaluating prediction bias, but they do not evaluate the models' uncertainty quantification. For this reason we introduce two additional evaluation scores: the continuous ranked probability score (CRPS) and the 90 % coverage. The CRPS is in general given by

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

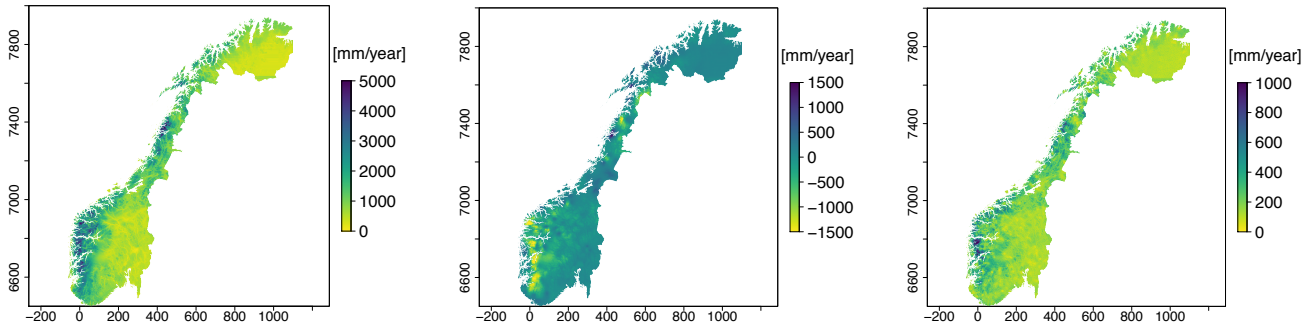
where y is the observed value and $F(\cdot)$ is the predictive cumulative distribution (Gneiting and Raftery, 2007). The CRPS takes the whole posterior distribution $F(\cdot)$ into account, unlike RMSE, ANE and NSE that only consider point predictions. A low CRPS corresponds to an accurate prediction, and the CRPS increases if the observed value y falls outside the posterior predictive distribution $F(\cdot)$. In this application, we assume $F(\cdot)$ to be Gaussian distributed with expected value given by the predicted mean annual runoff and standard deviation equal to the corresponding predictive standard deviation. The Gaussian assumption should be reasonable, as the posterior distributions of the predicted runoff typically are symmetric with light tails. We use the average CRPS over the 127 fully gauged catchments as a summary score.

The 90 % coverage is defined as the probability that 90 % of the observed values are covered by the corresponding 90 % posterior prediction intervals. This probability is computed empirically based on the predictions for the 127 fully gauged catchments, assuming that the SVC and Top-Kriging predictions follow a Gaussian distribution. If the empirical probability is close to 0.9, it suggests that the model provides an appropriate uncertainty quantification.

6 Results

6.1 Gridded mean annual runoff map for 1981-2010

In Figure 5a we present the runoff map produced by the $\mathrm{SVC_{PP}}$ approach. The difference between the new map and the original HBV product is visualized in Figure 5b, while the map's uncertainty estimates are shown in Figure 5c. Figure 5b shows that the $\mathrm{SVC_{PP}}$ map gives lower values of mean annual runoff in western Norway compared to the original HBV map. The difference is around 700-1500 mm/year. In eastern Norway, the original HBV map and the $\mathrm{SVC_{PP}}$ maps are approximately equal, both in



(a) Posterior mean SVC_{PP}. (b) Difference (SVC_{PP} minus HBV). (c) Posterior standard deviation SVC_{PP}.

Figure 5. Posterior mean of $q(\boldsymbol{u})$ for all grid nodes $\boldsymbol{u}$, difference between the new map and the original HBV map and posterior standard deviation of $q(\boldsymbol{u})$.

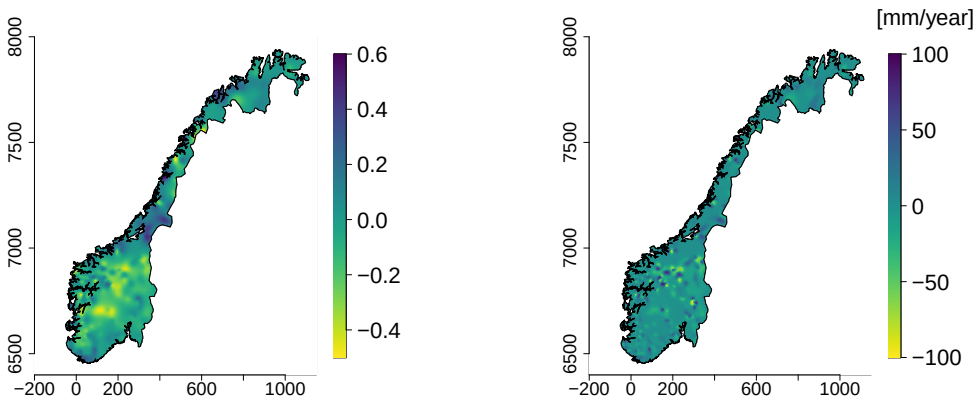
Table 1. Posterior median (0.025 quantile , 0.975 quantile) for the parameters of the SVC_{PP} model.

Parameter [unit]	SVC_{PP}
β_0 [mm/year]	153 (110,196)
β_1 [1]	0.83 (0.78,0.90)
ρ_x [km]	10.7 (5.4,26.1)
σ_x [mm/year]	117 (33.8,292)
ρ_α [km]	39.2 (29.4,51.9)
σ_α [1]	0.24 (0.21,0.27)
σ_y [mm/year]	205 (177,1000)

south-east and north-east. Around the glacier *Svartisen*, located in northern Norway in the area where Norway is quite narrow, the mean annual runoff of the SVC_{PP} map is lower than the mean annual runoff of the original HBV map. The difference here is around 1500 mm/year.

We see from Figure 5a that the SVC_{PP} map preserves most of the details provided by the original gridded HBV product in Figure 2a. The runoff map produced by SVC_{PP} also looks visually good without e.g. unrealistic jumps or obvious discontinuities. One exception is a line or discontinuity close to the Finnish border, north-east in Figure 5a, but this line was already present in the original HBV product in Figure 2a.

The covariate $h(\boldsymbol{u})$ makes a large contribution to the final model with a regression coefficient β_1 that is estimated to be 0.83. This can be seen in Table 1 where we present the parameter estimates of the SVC model. In Table 1 we also see that the marginal standard deviations σ_α and σ_x of the two spatial fields $\alpha(\boldsymbol{u})$ and $x(\boldsymbol{u})$ are of considerable magnitude, confirming that there indeed is a regional trend in the fit between the original HBV product and the actually observed mean annual runoff. The regional trend can be studied in Figure 6 where we show a visualization of the two spatial fields $\alpha(\boldsymbol{u})$ and $x(\boldsymbol{u})$. We see that



(a) Spatial component $\alpha(\boldsymbol{u})$ (unitless).

(b) Spatial component $x(\boldsymbol{u})$ (mm/year).

Figure 6. Posterior means for the two GRFs $x(\boldsymbol{u})$ and $\alpha(\boldsymbol{u})$ for SVC_{PP} for the Norwegian mainland.

the spatial pattern in Figure 5a mostly originates from the spatially varying coefficient component $\alpha(\boldsymbol{u})$ for SVC_{PP} (Figure 6a). The other GRF $x(\boldsymbol{u})$ contributes with more local adjustments in the mean annual runoff (Figure 6b). The spatial fields have hence picked up both long ranged and short ranged processes.

The posterior standard deviation of the SVC_{PP} in Figure 5c shows two trends: (i) The posterior uncertainty follows the pattern we see in the original HBV map in Figure 2a and (ii) if we look closely, the uncertainty is decreased in areas where there are observations, particularly around the centroids of the gauged catchments. Here, it is the spatially varying coefficient $(\beta_1 + \alpha(\boldsymbol{u})) \cdot h(\boldsymbol{u})$ from Equation (5) that causes pattern (i), while only including the GRF $x(\boldsymbol{u})$ would give pattern (ii). The component $\alpha(\boldsymbol{u})h(\boldsymbol{u})$ this way allows for a variance that is inhomogeneous in space given the process-based product $h(\boldsymbol{u})$. Figure 5c further shows that the SVC model gives quite high posterior standard deviations in a small area in western Norway, south of *Sognefjorden*. This can be explained by that this both is an area where we have few observations (see Figure 1b) and where the original HBV map performs poorly and overestimates the true runoff.

In Figure 7 we present a scatter plot that shows the fit between the runoff map in Figure 5a and the observed catchment mean annual runoff. The results show that the SVC_{PP} map corresponds considerably better with the observed runoff for the fully gauged catchments than the original HBV map (Figure 2b). The original HBV map gave a correlation of 0.933 between the predictions and the observations for the fully gauged catchments, while the corrected SVC_{PP} map gives a correlation approximately equal to 1.

We also investigated the correlation between the map and the observed runoff for the partially gauged catchments where we only have 1-29 years of measurements in the 30 year time period of interest (Figure 7). For these catchments, the original HBV model gave a correlation of 0.917. The SVC_{PP} map gives correlation 0.986. The correlations and Figure 7 indicate that the SVC_{PP} map provides a better fit for the partially gauged catchments than the original HBV map. Here, we can not be entirely sure because the underlying observations from the partially gauged catchments in Figure 7 only are approximations of the true

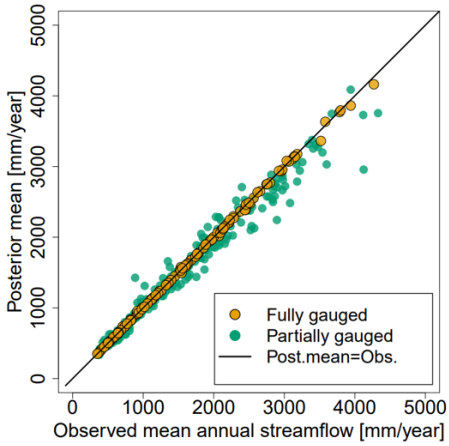


Figure 7. Scatter plot showing the predicted mean annual runoff for SVC_{PP} and the observed mean annual runoff from fully gauged (orange) and partially gauged catchments (green).

Table 2. Predictive performance for the cross-validation experiments when the target catchments are treated as ungauged (UG) and partially gauged (PG) for the HBV model, the suggested SVC model and for Top-Kriging (TK). Recall that subscript PP refer to the geostatistical preprocessing step. The results from the geostatistical preprocessing method (PP) are also included as a reference (without any further analysis) for a better understanding of the other results. The best method for each evaluation criterion is marked in bold.

		Ungauged target (UG)			Partially gauged target (PG)		
	HBV	SVC_{PP}	TK_{PP}	PP	SVC_{PP}	TK_{PP}	PP
RMSE (mm/yr)	394	**315**	350	389	166	181	**134**
ANE	0.180	**0.111**	0.125	0.192	0.054	0.053	**0.047**
NSE	0.815	**0.881**	0.854	0.771	0.968	0.961	**0.978**
CRPS (mm/yr)	235	**145**	173	209	73	77	**71**
Coverage (90 %)	×	0.83	**0.91**	0.96	0.95	**0.94**	1

runoff between 1981-2010, computed based on 1-29 annual observations from this time period. It is however a good sign that the fit for the partially gauged catchments (green) is not as good as for the fully gauged catchments (orange).

6.2 Cross-validation for ungauged and partially gauged catchments

In Table 2 we present the results from the cross-validation assessment described in Section 5.2. For ungauged catchments (UG), we find that the RMSE of our SVC_{PP} method is 20 % lower than the RMSE of the HBV model. Compared to Top-Kriging, the SVC_{PP} model gives 10 % lower RMSE. The ranking between the models is the same also for the ANE, NSE and CRPS. When it comes to uncertainty quantification, Top-Kriging gives the best uncertainty representation for ungauged catchments according

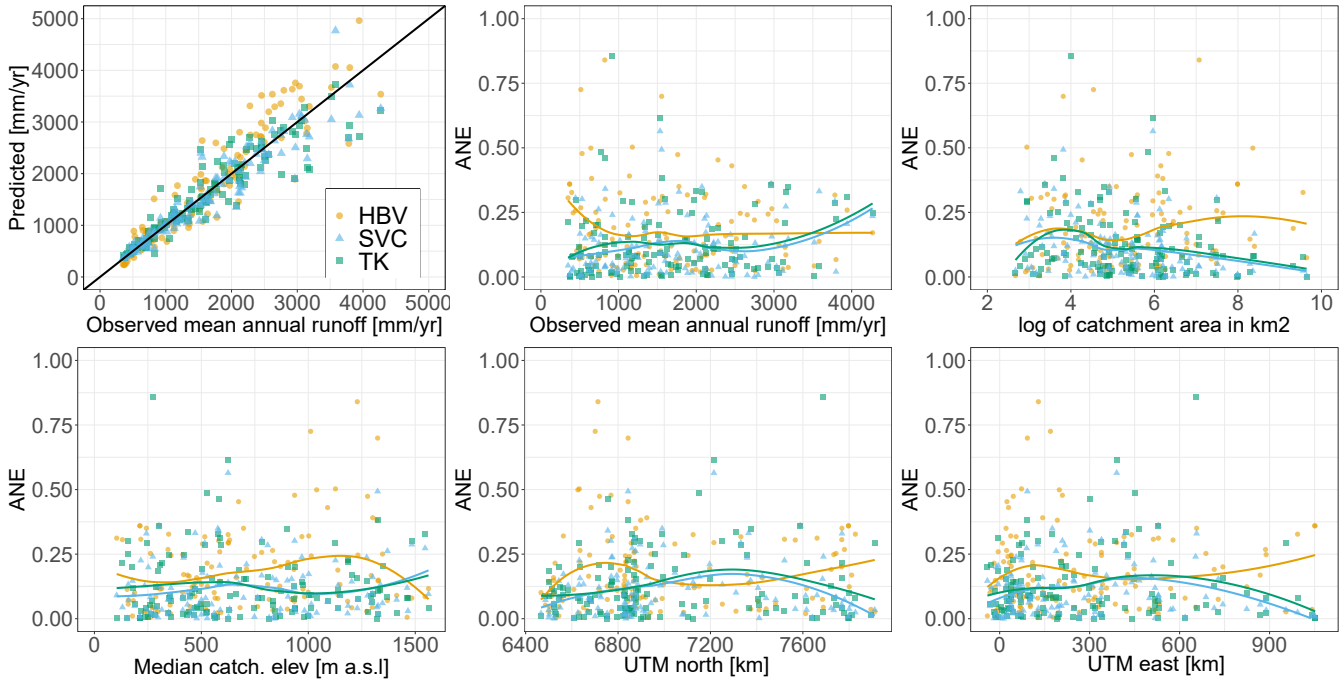


Figure 8. Predictive performance of the methods (HBV, SVC_{PP} and TK_{PP}) for predictions in ungauged catchments (UG) performed by cross-validation. The first plot shows the fit between the predictions and the observations for the methods. The remaining plots show the ANE for each of the 127 cross-validation catchments plotted against some selected catchment attributes; more specifically the observed runoff, catchment area, median catchment elevation, UTM33 north and UTM33 east. The fitted curves are regression splines (made by `geom_smooth()` in `R`).

to the 90 % coverage, with 0.91 coverage. However, SVC_{PP} also performs acceptable with 0.83 coverage on a cross-validation performed on (only) 127 catchments.

In Table A1 in the Appendix, we include the methods' predictive performance for ungauged catchments when not using the preprocessing step and short records (SVC and TK). These results give the same ranking between the methods as before, but with one exception. SVC performs approximately as good as Top-Kriging in terms of 90% coverage, with coverages of 0.87 and 0.94 respectively. From Table A1 we also notice that the difference in performance between the SVC model and Top-Kriging is larger for this setting, where we had fewer observations. This is reasonable as we can expect the SVC model to be more robust than a purely data-driven model if the data availability is poorer.

Further, we compared the predictive performance for ungauged catchments (UG) for the SVC_{PP} approach, the HBV model and Top-Kriging (TK_{PP}) across the study area and across catchment attributes in terms of the absolute normalized error (ANE). This is visualized for some selected catchment attributes in Figure 8. We see that the HBV model in general tends to over-estimate the mean annual runoff. It gives the highest ANE values in the south-western part of the country, and particularly for catchments at higher elevations (800-1400 m a.s.l). The latter might be due to the interpolated precipitation product used

as input in the HBV-model, where orographic enhancement of precipitation is accounted for by an elevation gradient. Since precipitation gauging stations seldom are located at high elevations (Lussana et al., 2018), the precipitation is extrapolated to the highest altitudes giving rise to biases in the precipitation field. Figure 8 further shows that the two geostatistical approaches (SVC_{PP} and TK_{PP}) perform better than the HBV model for catchments with mean elevations in the range 800-1400 m a.s.l. This demonstrates that the SVC approach is able to compensate for its poor HBV input in these areas.

The lines in Figure 8 next show that Top-Kriging and SVC_{PP} in general tend to follow the same trends across catchments attributes. For example, both perform well for catchments with large drainage areas, supporting existing results from Viglione et al. (2013) regarding the predictive performance of Top-Kriging. For catchments with large drainage areas, there are typically data from overlapping subcatchments available which makes areal referenced geostatistical models particularly appropriate. The two geostatistical approaches also perform well for catchments located in the eastern parts of Norway. In the south-eastern Norway we find catchments with larger drainage areas and most of them are located at relatively low elevations. The data availability is also good in the south-eastern parts of Norway, suitable for geostatistical modeling. It is hard to see a clear trend in when SVC_{PP} performs better than Top-Kriging from Figure 8, but we notice that Top-Kriging (and the HBV model) in general produce more extreme ANE values than the SVC model.

We next consider the performance of the models for predictions in partially gauged (PG) catchments. The results in Table 2 show that we obtain a large reduction in the predictive performance for the SVC_{PP},PG case compared to the case when we have no data from the target catchments (SVC_{PP},UG). The reduction in RMSE is 47 $\%$. The improvement for SVC_{PP},PG compared to the HBV model is 58%. Compared to Top-Kriging, the SVC_{PP} approach is slightly better in terms of RMSE, but approximately equally good in terms of ANE, NSE, CRPS and 90 $\%$ coverage. Table 2 also shows that the Top-Kriging estimates are substantially improved when including preprocessed short records from the target catchments in the likelihood (PG compared to UG for TK_{PP}).

The improved performance of TK_{PP} and SVC_{PP} for the PG case is mainly caused by the preprocessing procedure's ability to perform (very) accurate predictions of Norwegian mean annual runoff when a few annual observations are available. In 2 we see that the input data provided by the preprocessing step (PP) alone gives predictions that are better than the predictions of the SVC_{PP} and TK_{PP} approaches. The improved results for TK_{PP} and SVC_{PP} however, show that the two geostatistical methods are able to exploit the good performance of PP, and that the SVC approach indeed can be used to combine both process-based data and data from fully gauged and partially gauged catchments.

7 Discussion

We have presented a geostatistical model for mean annual runoff that incorporates simulations from a process-based model through a spatially varying coefficient and shown how short records can be included by using the methodology from Roksvåg et al. (2020) for filling in missing values.

In a preliminary study we tested models with only one spatial field, i.e. only either $x(\boldsymbol{u})$ or $\alpha(\boldsymbol{u})$ was included in Equation (5). These models performed quite well in terms of both posterior mean and posterior uncertainty for the Norwegian dataset,

which indicates that it might be satisfactory to use a model with only one spatial field for many study areas. However, our preliminary experiments also showed that a model with two spatial fields often gave a more more realistic spatial distribution of uncertainty than a model with only one spatial field. Further, Figure 6 showed that the model was able to capture both short and long ranged processes through its two fields, which can be a useful model property that can avoid that the model smooths out the process-based covariate too much. In general, the importance of $x(\boldsymbol{u})$ compared to $\alpha(\boldsymbol{u})$ depends on the study area, the data availability and the quality of the process-based input model.

Table 2 showed that TK_{PP} and SVC_{PP} performed slightly poorer than the preprocessing input model alone for the partially gauged catchments (PG). When constructing the models, we did not want the SVC approach and Top-Kriging to put too much weight on the more uncertain preprocessed short records. The latter was included in the model by specifying a larger (prior) observation uncertainty for the partially gauged catchments (0-23 % of the observed value) compared to the fully gauged catchments (0-6 % of the observed value). We have not tested how this uncertainty specification affects the results, but in future work, the SVC_{PP} model and TK_{PP} could be improved by selecting the observation uncertainty for the preprocessed data more carefully. The observation uncertainty for the partially gauged catchments can e.g. be set independently of the fully gauged catchments and based on the record length of the short records. An option could also be to use the predictive uncertainty of the preprocessing method to specify the (prior) measurement uncertainty for the partially gauged catchments in the SVC model and Top-Kriging.

In the article, we presented a framework for estimating mean annual runoff, which is one of several key flow indices. The SVC framework can be used for other flow indices as well, but the computational complexity makes it most suitable for flow indices of longer temporal scale or for modeling long-term averages. In Equation (7), we also assume a linear aggregation of runoff which is particularly appropriate for mass conservative variables like annual runoff. If the modeler want to avoid this assumption, two simple model modifications are possible:

1) Make the runoff observations point referenced by letting $Q(\mathcal{A}) = q(\boldsymbol{u}_{\mathcal{A}})$ in Equation (7), where $\boldsymbol{u}_{\mathcal{A}}$ is the centroid of catchment $\mathcal{A}$ and $q(\cdot)$ is point runoff as defined in Equation (5). This modification also allows for doing inference on log transformed data.

2) Add more covariates or random noise outside the integral in Equation (7). This way the areal representation is preserved, but it becomes easier to violate the water balance constraints.

A potential weakness of the proposed model is that it consists of Gaussian components which can result in negative runoff estimates. Negative estimates can occur if the flow index and the corresponding study area have runoff observations close to zero. This is another argument for using the SVC model mainly for flow indices of a longer temporal scale. Negative predictions can also occur if the discretization of the study area and/or the SPDE mesh is too coarse relative to the spatial variability of the target variable. In the study presented here, no negative values were produced.

Figure 7 showed that the SVC_{PP} gave a very good fit for the 127 fully gauged catchments, almost entirely reproducing the actual observed mean annual runoff in the resulting gridded map. We emphasize that the proposed method is not guaranteed to reproduce the observed value with the precision we saw in this case study. How good the fit is for the fully gauged catchments depends on the data quality, the gauging density and the complexity of the spatial variability of the underlying hydrological

process. Obtaining a correlation around 1 for the gauged catchments, as in Figure 7, is not necessarily desirable either, as it might affect the fit for the ungauged catchments negatively. This might explain the over-confidence of the SVC_{PP} model, expressed through the 83 % coverage for the UG case, in Table 2. It is possible to influence the model fit by making the prior observation uncertainty of $s_i \cdot \sigma_y^2$ wider or narrower.

In Norway the gauging density is moderate. We expect the suggested SVC model to outperform purely geostatistical methods like Top-Kriging for gauging densities that are low to moderate. For data sparse areas, the process-based information provided by the HBV model is probably more important. This claim is based on intuition about the models under discussion, but is also indicated by our results. Top-Kriging is closer to the SVC model in predictive performance for the dataset where we use data from 411 catchments (UG in Table 2) than for the reduced dataset where we only use data from 127 catchments (Table A1 in the Appendix).

Whether the suggested framework performs better than a purely geostatistical method is of course also connected to the quality of the process-based input model and the calibration procedures performed on the hydrological product. However, our results have clearly demonstrated that it is possible to improve a process-based hydrological product by using the suggested framework. All experiments showed that the SVC approach improved the predictions compared to the original HBV simulations. The SVC model can hence be considered as an objective approach for correcting the simulations from a process-based model, and reduce the need for subjective, manual corrections.

8 Conclusions

We have presented a Bayesian geostatistical model for annual runoff estimation that incorporates simulations from a process-based hydrological through a covariate whose regression coefficient is allowed to vary in the study area according to a Gaussian random field. A preprocessing step for including short records was also suggested such that the model could exploit data from both fully gauged and partially gauged catchments.

The model was evaluated by predicting mean annual runoff data for Norway (1981-2010), and simulations from the process-based HBV model were used as a covariate. The results showed that the suggested framework outperforms a purely process-based model when predicting runoff in ungauged and partially gauged catchments. The reduction in RMSE was 20 % for ungauged catchments and 58 % for partially gauged catchments. The increased predictive performance obtained compared to a purely process-based model is connected to the quality of the process-based product and the calibration procedures performed on it. However, all results show that the suggested framework is able to improve the predictions from a process-based model. The approach can hence be used as a objective method for correcting process-based runoff maps relative to data. The large reduction in RMSE for partially gauged catchments demonstrates that the preprocessing method from Roksvåg et al. (2020) can be incorporated into the proposed model to exploit short records.

Furthermore, the suggested model gave a 10 % lower RMSE than a purely geostatistical method (Top-Kriging) when predicting runoff in ungauged catchments. Particularly if the gauging density is low to moderate, the suggested framework is expected to outperform purely geostatistical models. For catchments that had a few annual streamflow observations available,

a purely geostatistical method performed equally well (Top-Kriging) or slightly better (PP) than the proposed approach. Since most study areas consist of a mix of ungauged, fully gauged and partially gauged catchments, the proposed SVC model stands out as a good approach for making a consistent gridded runoff map for a larger area.

Author contributions. TR: Did the experiments, wrote R code, wrote the majority of the paper and made figures. IS: Came up with initial ideas, contributed with discussion throughout the work and with ideas for experimental set-up. KE: Provided the data, contributed with discussion throughout the work, particularly regarding the HBV model, the data and interpretation of results. Contributed to the writing of Chapter 2.

Competing interests. No competing interests are present.

Code and data availability. The data and code used in this study can be provided by the main author upon request.

Acknowledgements. We would like to thank Stein Beldring for providing the gridded HBV product and for valuable discussions around the work. We would also like to thank Erik Holmqvist for discussions about the data. The project was funded by the Reasearch council of Norway, grant number 250362.

References

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., and Lindgren, F.: Spatial modeling with R-INLA: A review, WIREs Computational Statistics, 10, https://doi.org/10.1002/wics.1443, 2018.

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.

Beldring, S., Roald, L. A., and Voksø, A.: Arenningskart for Norge. Årsmiddelverdier for avrenning 1961-1990., Tech. Rep. Oslo: NVE, http://publikasjoner.nve.no/dokument/2002/dokument2002_02.pdf, 2002.

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

Bergström, S.: Development and Application of a Conceptual Runoff Model for Scandinavian Catchments, vol. 134 pp., 1976.

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

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.

Borgvang, S., Stålnacke, P., Skarbøvik, E., Beldring, S., Selvik, J., Tjomsland, T., and Harsten, S.: Riverine inputs and direct discharges to Norwegian coastal waters -2004. OSPAR Commission for the Protection of the Marine Environment of the North-East Atlantic., Norwegian Institute for Water Research, Report No. 5135-2006, 159 pp., 2006.

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.

Clark, M., Bierkens, M., Samaniego, L., Woods, R., Uijlenhoet, R., Bennett, K., Pauwels, V., Cai, X., Wood, A., and Peters-Lidard, C.: The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism, Hydrology and Earth System Sciences Discussions, pp. 1–14, https://doi.org/10.5194/hess-2016-693, 2017.

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

Doherty, J.: PEST: Model Independent Parameter Estimation. Fifth edition of user manual., Watermark Numerical Computing, Brisbane, Australia, 2004.

Fatichi, S., Vivoni, E., Ogden, F., Ivanov, V., Mirus, B., Gochis, D., Downer, C., Camporese, M., Davison, J., Ebel, B., Jones, N., Kim, J., Mascaro, G., Niswonger, R., Restrepo, P., Rigon, R., Shen, C., Sulis, M., and Tarboton, D.: An overview of current applications, challenges, and future trends in distributed process-based models in hydrology, Journal of Hydrology, 537, 45 – 60, https://doi.org/https://doi.org/10.1016/j.jhydrol.2016.03.026, 2016.

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

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

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H.: Constructing Priors that Penalize the Complexity of Gaussian Random Fields, Journal of the American Statistical Association, 114, 445–452, https://doi.org/10.1080/01621459.2017.1415907, 2019.

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

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

Gelfand, A., Hyon-Jung, K., Sirmans, C., and Banerjee, S.: Spatial Modeling With Spatially Varying Coefficient Processes, Journal of the American Statistical Association, 98, 387–396, https://doi.org/10.2307/30045248, 2003.

Gelfand, A., Diggle, P., Fuentes, M., and Guttorp, P.: Handbook of Spatial Statistics, Chapman & Hall, https://doi.org/10.1201/9781420072884, 2010.

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

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

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

Guttorp, P. and Gneiting, T.: Studies in the history of probability and statistics XLIX On the Matérn correlation family, Biometrika, 93, 989–995, https://doi.org/10.1093/biomet/93.4.989, 2006.

Hanssen-Bauer, I., Førland, E. J., Haddeland, I., Hisdal, H., Mayer, S., Nesje, A. ., and Sorteberg, A.: Climate in Norway 2100 - a knowledge base for climate adaption., NCCS Report 1/2017. Retrieved from Oslo, Norway., 2017.

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

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

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

Laaha, G., Skøien, J., Nobilis, F., and Blöschl, G.: Spatial Prediction of Stream Temperatures Using Top-Kriging with an External Drift, Environmental Modeling & Assessment, 18, https://doi.org/10.1007/s10666-013-9373-3, 2013.

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

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, https://doi.org/10.1111/j.1467-9868.2011.00777.x, 2011.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, M.: Development and test of the distributed HBV-96 hydrological model, Journal of Hydrology, 201, 272 – 288, https://doi.org/https://doi.org/10.1016/S0022-1694(97)00041-3, 1997.

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

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., and Andersen, J.: seNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day, Earth System Science Data, 10, 235–249, https://doi.org/10.5194/essd-10-235-2018, https://essd.copernicus.org/articles/10/235/2018/, 2018.

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

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

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

Pannecoucke, L., Le Coz, M., Freulon, X., and de Fouquet, C.: Combining geostatistics and simulations of flow and transport to characterize contamination within the unsaturated zone, Science of The Total Environment, 699, 134 216, https://doi.org/https://doi.org/10.1016/j.scitotenv.2019.134216, http://www.sciencedirect.com/science/article/pii/S0048969719341932, 2020.

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

Roksvåg, T., Steinsland, I., and Engeland, K.: Estimation of annual runoff by exploiting long-term spatial patterns and short records within a geostatistical framework, Hydrology and Earth System Sciences, 24, 4109–4133, https://doi.org/10.5194/hess-24-4109-2020, https://hess.copernicus.org/articles/24/4109/2020/, 2020.

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.

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

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

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, https://doi.org/10.1080/02626660009492385, 2000.

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, https://doi.org/10.1214/16-STS576, 2017.

Skarbøvik, E., Stålnacke, P., Kaste, ø., Selvik, J., Tjomsland, T., Høgåsen, T., Aakerøy, P., Haaland, S., and Beldring, S.: Riverine inputs and direct discharges to Norwegian coastal waters -2008., Norwegian Pollution Control Authority, Report no. TA-2569-2009,75 pp., 2009.

Skøien, J.O.: Package 'rtop', https://cran.r-project.org/web/packages/rtop/rtop.pdf, [Online; accessed 25-October-2021], 2018.

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

Sælthun, N.: The Nordic HBV model., Technical Report 07. Norwegian Water Resources and Energy Directorate., http://folk.uio.no/nilsroar/gf247/hbvmod.pdf, 1996.

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

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, https://doi.org/10.1029/2007JD009006, 2008.

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

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, https://doi.org/10.5194/hessd-10-449-2013, 2013.

Wang, Y., Akeju, O. V., and Zhao, T.: Interpolation of spatially varying but sparsely measured geo-data: A comparative study, Engineering Geology, 231, 200 – 217, https://doi.org/https://doi.org/10.1016/j.enggeo.2017.10.019, 2017.

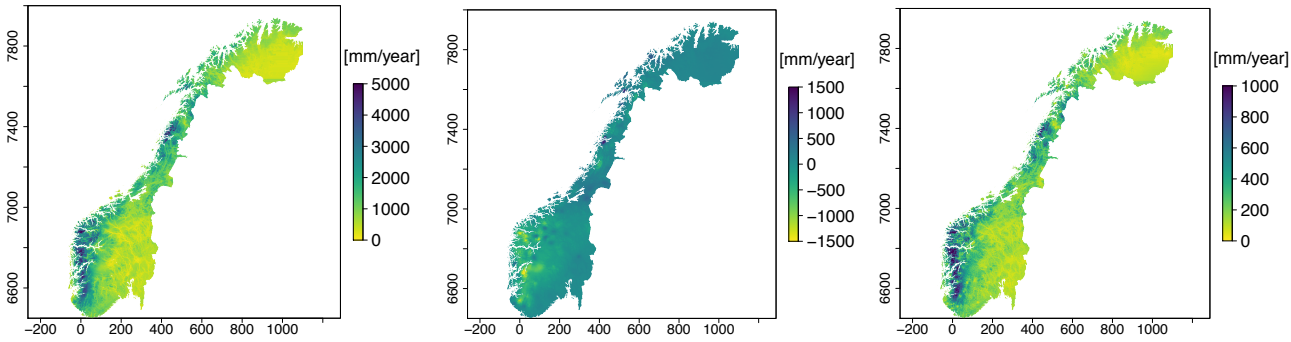
Whittle, P.: On stationary processes in the plane, Biometrika, 41, 434–49, 1954.

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

WMO: International meteorological vocabulary, 1992.

Appendix A: Results when omitting short records

We repeat the experiments from Section 5.1 and 5.2 for ungauged catchments, but we only use observations from the 127 fully gauged catchments in Figure 1a. The runoff data from the partially gauged catchments are simply removed from the analysis. The experiments are included to show that the SVC model works regardless of preprocessing.



(a) Posterior mean SVC. (b) Difference (SVC minus HBV). (c) Posterior standard deviation SVC.

Figure A1. Posterior mean of $q(\boldsymbol{u})$ for all grid nodes, difference between the new map and the original HBV map and posterior standard deviation of $q(\boldsymbol{u})$. The model is fitted without including short records.

The runoff map provided by the SVC model, when not using short records, is shown in Figure A1. The maps look similar to the maps in Figure 5, but the posterior uncertainty is larger in western Norway in Figure A1c. The reasons are that there are less observations available from western Norway in the dataset consisting only of fully gauged catchments and that this is an area with large deviance between the original HBV map and the observed streamflow.

In Figure A2 we show the fit between the observed runoff and the runoff predicted by the map in Figure A1a. The fit is good for the fully gauged catchments, as before. The fit is also improved for the partially gauged catchments compared to the

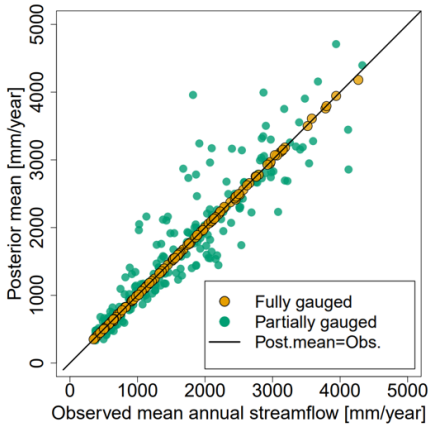


Figure A2. Scatter plot showing the predicted mean annual runoff (posterior mean of $Q(\mathcal{A})$) for SVC and the observed streamflow from fully gauged and partially gauged catchments when short records are omitted from the likelihood.

Table A1. Predictive performance for cross-validation when the target catchments are treated as ungauged (UG) for the HBV model, the suggested SVC model and for Top-Kriging (TK). Short records are omitted from the observation likelihood and the preprocessing step is not performed. The best method for each evaluation criterion is marked in bold.

	UG		
	HBV	SVC	TK
RMSE (mm/yr)	394	**320**	381
ANE	0.180	**0.135**	0.176
NSE	0.815	**0.878**	0.827
CRPS (mm/yr)	235	**156**	211
Coverage (90 %)	×	**0.87**	0.94

original HBV map in Figure 2a. Here, the original HBV model gave correlation 0.917 between observed and predicted values, while the map in Figure A1a gives correlation 0.924. However, when short records and preprocessing were included in the analysis, the correlation was 0.986 (SVC_{PP} in Figure 7). This illustrates the reduced predictive performance when omitting short records from the analysis in Norway and in countries with similar spatio-temporal trends in annual runoff.

The cross-validation results for the experiments where we omit the short records are summarized in Table A1. Again the SVC model performs considerably better than the HBV model and Top-Kriging in terms of RMSE, ANE, NSE and CRPS. Remark that the difference in performance from Top-Kriging is larger for this dataset (for UG), compared to when using the larger dataset (Table 2).