A Bayesian geostatistical model that incorporates the simulations from a process-based hydrological model through a spatially varying coefficient (SVC)

Thea Roksvåg (1), Ingelin Steinsland (1) & Kolbjørn Engeland (2)

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

May 2020

Motivation

Main goal: Predict mean annual runoff in ungauged basins.

Working hypothesis 1: A geostatistical model that is able to incorporate the simulations from a process-based hydrological model will perform better than a purely geostatistical method or a purely process-based method.

Working hypothesis 2: The potential information stored in all available streamflow observations should be exploited, also information from short records.

Nærøydalen valley, Norway (https://blog.historichotelsofeurope.com/hiking-in-europe/voss-norway/)

Suggested approach: A Bayesian SVC model

The true mean annual runoff at a point location $\boldsymbol{u}$ is given by:

$$q(\boldsymbol{u}) = \beta_0 + (\beta_1 + \alpha(\boldsymbol{u})) \cdot HM(\boldsymbol{u}) + x(\boldsymbol{u}); \qquad \boldsymbol{u} \in \mathcal{R}^2$$
$$x(\boldsymbol{u}) \sim \mathrm{GRF}(\sigma_x, \rho_x)$$
$$\alpha(\boldsymbol{u}) \sim \mathrm{GRF}(\sigma_\alpha, \rho_\alpha)$$

$HM(\boldsymbol{u})$ is a simulation produced by a process-based hydrological model on a grid cell $\boldsymbol{u}$. It is treated as a covariate.

β_0 is an intercept and β_1 is a fixed regression coefficient.

$x(\boldsymbol{u})$ and $\alpha(\boldsymbol{u})$ are Gaussian random fields (spatial components).

The relationship between the true runoff (response variable) and the simulations from the hydrological model (covariate) is allowed to vary in space through the spatially varying coefficient (SVC). The dependency structure of the relationship follows the GRF $\alpha(\boldsymbol{u})$.

Observation model

Mean annual runoff is observed for catchment areas [mm/year]. The observation model is:

$$y_i = \frac{1}{n_i} \sum_{\boldsymbol{u} \in \mathcal{L}_{\mathcal{A}_i}} q(\boldsymbol{u}) + \epsilon_i$$

$$\epsilon_i \sim \mathcal{N}(0, \sigma_y^2).$$

$\mathcal{L}_{\mathcal{A}_i}$ is a discretization of catchment $\mathcal{A}_i$ with n_i grid nodes.

We specify prior distributions for all model parameters $\beta_0, \beta_1, \sigma_y, \rho_x, \sigma_x, \rho_\alpha$ and σ_α.

Approximative Bayesian inference is achieved by using INLA and the SPDE approach to spatial modeling:

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series B, 2009.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society: Series B, 2011.

Exploiting short records (optional)

To exploit short records of data, we perform record augmentation for partially gauged catchments as a pre-processing step.

The approach from Roksvåg et al. (2019) is used.

This works well for areas that are driven by repeated runoff patterns over time. For example:

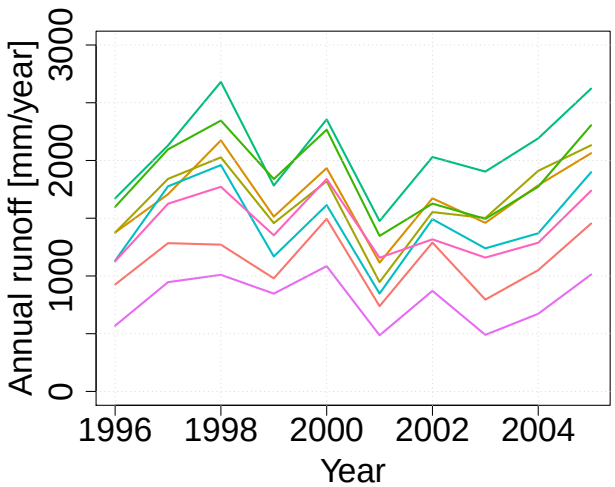
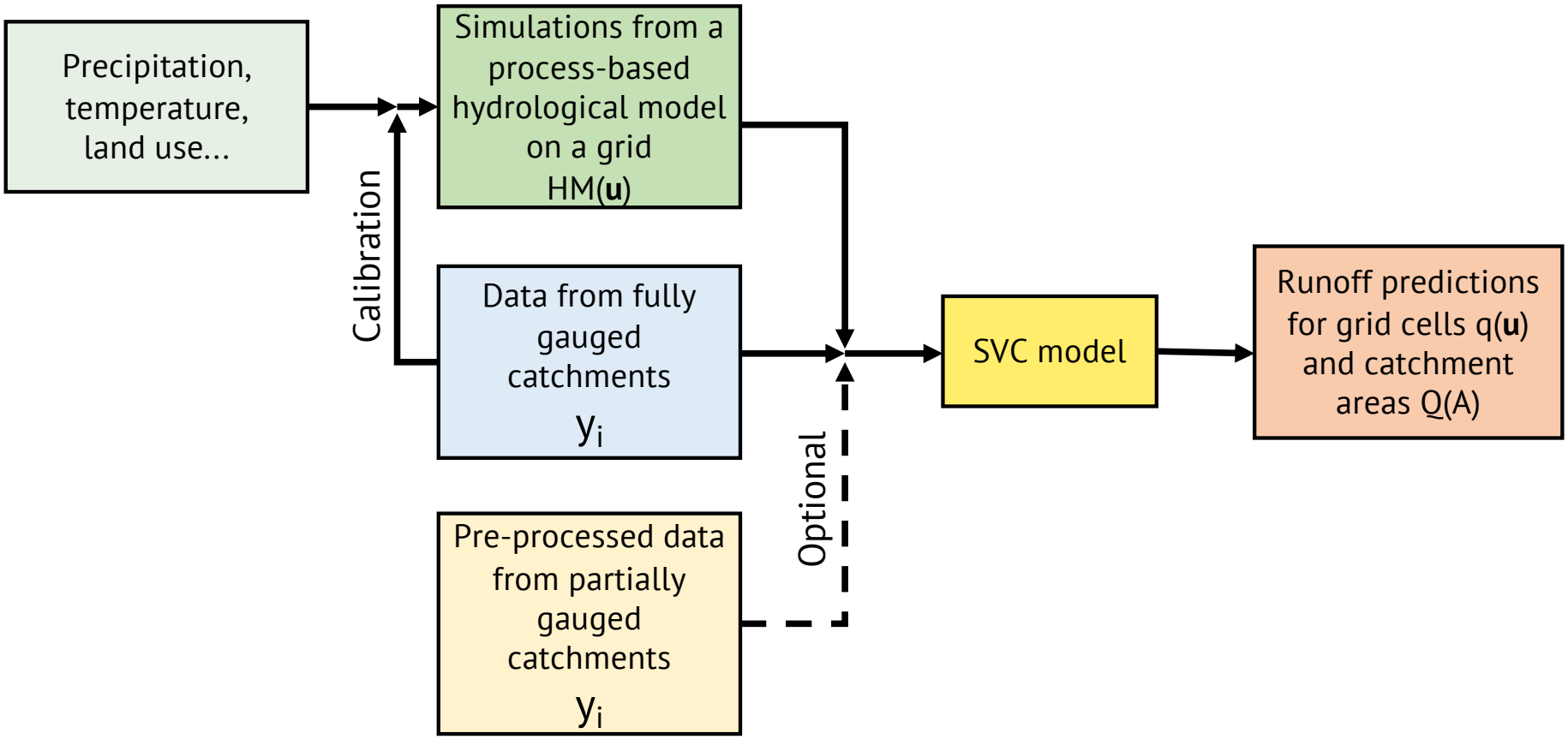


Fig: Time series of annual runoff from 8 catchments in Norway.

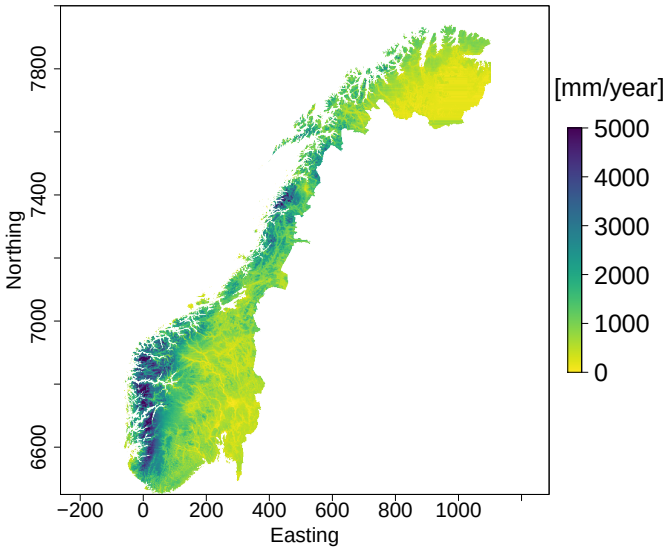
T. Roksvåg, I. Steinsland and K. Engeland. *A geostatistical framework for estimating flow indices by exploiting short records and long-term spatial averages – Application to annual and monthly runoff*, https://www.hydrol-earth-syst-sci-discuss.net/hess-2019-415/, 2019

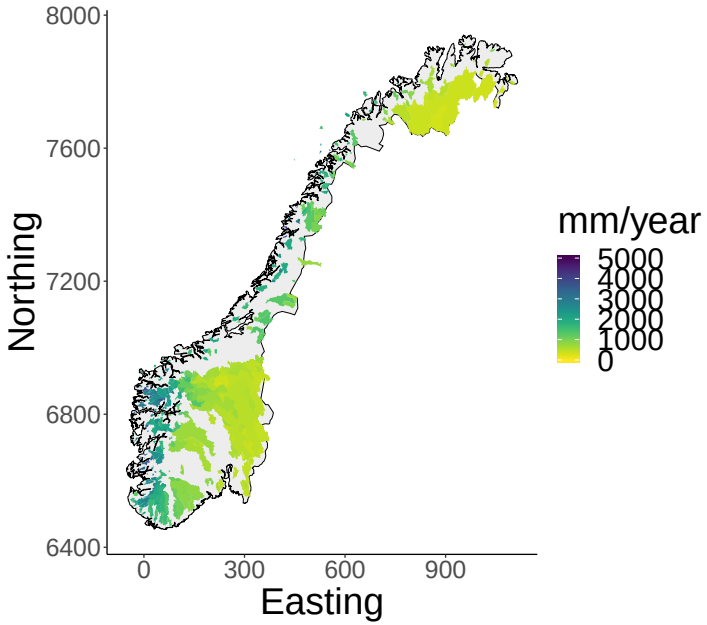
Workflow of the suggested approach



Model testing: Study area and input data

Mean annual runoff for Norway (1981-2010).



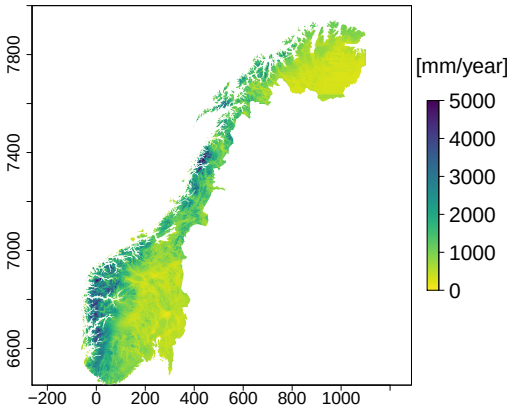


$HM(\boldsymbol{u})$: Simulations of mean annual runoff from the HBV model (Bergström, 1976) on a 1 km $\times$ 1 km grid.

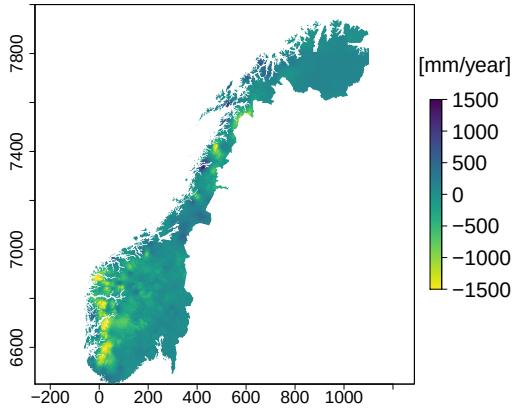
S. Bergström. Development and Application of a Conceptual Runoff Model for Scandinavian Catchments, volume 134 pp. 01 1976.

1) Streamflow observations from 127 fully gauged catchments.
2) Streamflow observations 284 partially gauged catchments that have at least one annual observation between 1981-2010.

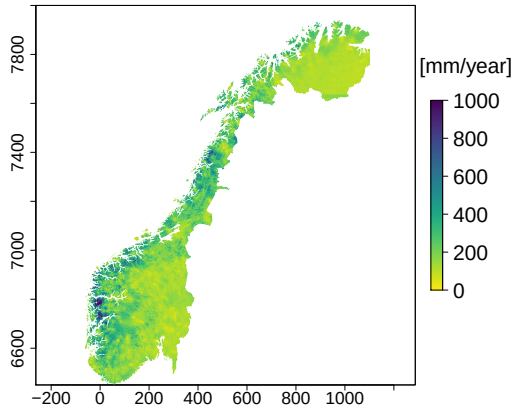
Resulting gridded mean annual runoff map



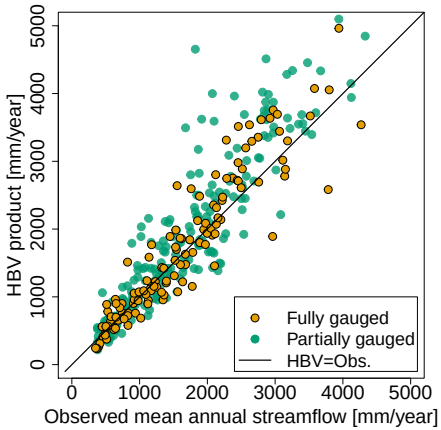
a) Posterior mean.



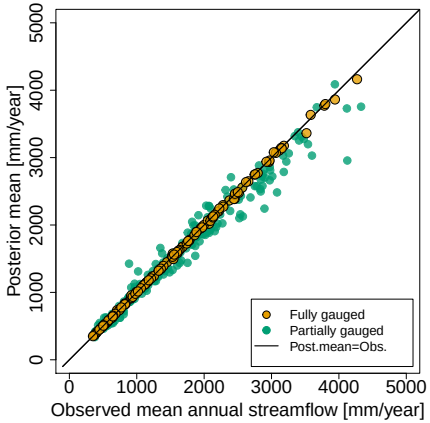
b) Difference from the original HBV map (SVC-HBV).



c) Posterior standard deviation.



Original HBV map



New SVC map

Cross-validation for 127 fully gauged catchments

The catchments are treated as:

a) Ungauged (UG) with zero observations between 1981-2010.

b) Partially gauged (PG) with three randomly drawn annual observations between 1981-2010.

The proposed SVC model is compared to a purely geostatistical method (GS) and the HBV model.

Subscript PP refer to that pre-processing of short records is performed before further analysis.

		Ungauged catch. (UG)		Partially gauged catch. (PG)	
Ev. score	HBV	SVC_{PP}	GS_{PP}	SVC_{PP}	GS_{PP}
RMSE [mm/yr]	394	**315**	353	166	**138**
ANE [1]	0.180	**0.111**	0.127	0.054	**0.047**
CRPS [mm/yr]	235	**145**	175	73	**65**

Main conclusions: (i) The SVC model performs better than the process-based model (HBV) and the geostatistical model (GS) for ungauged catchments.
(ii) The gain of including short records is large for Norwegian mean annual runoff.