Estimating mean annual runoff
by using a geostatistical spatially varying coefficient model that incorporates process-based simulations and short records

T.Roksvåg¹,², I. Steinsland² and K.Engeland³

¹ Norwegian Computing Center (NR), Oslo, Norway
² Norwegian University of Science and Technology (NTNU), Trondheim, Norway
³ The Norwegian Water Resources and Energy Directorate (NVE), Oslo, Norway

roksvag@nr.no
Motivation

Main goal: Estimate mean annual runoff in ungauged and partially gauged catchments.

How: We want to exploit several data sources and suggest to use a process-based model in combination with a geostatistical approach.
Data available for Norway (1981-2010)

1) Streamflow obs. from fully gauged catchments: 127 catchments with annual streamflow obs. from the whole 30 year target period.

Areal referenced data

2) Streamflow obs. from partially gauged catchments (short records): 284 catchments have annual streamflow obs. for 1-29 years in the 30 year target period.

Areal referenced data

3) Simulations of mean annual runoff from a process-based hydrological model

The HBV model [1] was used to make simulations on a 1 km x 1 km grid. Land use, temperature and precipitation information was used as input.

Gridded data
Geostatistical model

The true mean annual runoff \( q(u) \) at a location \( u \) is modeled as

\[
q(u) = \beta_0 + (\beta_1 + \alpha(u)) \cdot h(u) + x(u); \quad u \in \mathbb{R}^2
\]

\[
x(u) \sim \text{GRF}(\sigma_x, \rho_x)
\]

\[
\alpha(u) \sim \text{GRF}(\sigma_\alpha, \rho_\alpha)
\]

\( h(u) \): The simulated mean annual runoff for grid cell \( u \), generated by a process-based model.

\( \text{GRF}(\sigma, \rho) \): Gaussian random field with range \( \rho \) and standard deviation \( \sigma \) that captures spatial variability.

\( \beta_0 \) is an intercept and \( \beta_1 \) is a regression coefficient.

**Spatially varying coefficient (SVC):** The regression coefficient that models the relationship between the simulations \( h(u) \) and the true runoff \( q(u) \) varies in the study area according to \( \alpha(u) \).
Observation model

The true mean annual runoff $q(u)$ is observed through areal referenced streamflow observations, with uncertainty:

$$y_i = \frac{1}{n_i} \sum_{u \in L_{A_i}} q(u) + \epsilon_i \quad \epsilon_i \sim N(0, \sigma_y^2)$$

$L_{A_i}$ is a discretization of catchment $A_i$ with $n_i$ grid nodes.

The nested structure of catchments is taken into account by modeling the observations as areal referenced.
How to include short records?

We perform record augmentation as a preprocessing step for partially gauged catchments that have less than 30 years of data. This gives us $y_i$ for the partially gauged catchments.

The approach from Roksvåg et. al (2020) is used [2]. Here, we exploit that the Norwegian annual runoff follows repeated spatial patterns due to e.g. orographic precipitation.

The data $y_i$ from the fully gauged catchments are weighted more than the data from the partially gauged catchments in our observation likelihood.

Example: Time series of annual runoff from 8 Norwegian catchments. The ranking between the catchments, from wet to dry, is approximately constant over time. According to [2], this is a setting for which the value of short records is high.
The workflow of the SVC model

Statistical inference

Fast and approximative Bayesian inference by using INLA [3], and the SPDE [4] approach to spatial modeling
Results: A gridded mean annual runoff map

Posterior mean for SVC model

Difference between new SVC map and original HBV map (SVC – HBV)

Posterior standard deviation

The fit between the observations (x-axis) and the estimated runoff (y-axis) is considerably better for the new SVC map
Cross-validation for 127 catchments

We compare three methods:
SVC: The proposed approach.
GS: The geostatistical method from Roksvåg et. al. (2020) [2].
HBV: A process-based model.

Two evaluation settings:
Ungauged: The target catchments are treated as ungauged with zero annual observations between 1981 and 2010.
Partially gauged: The target catchments are treated as partially gauged and have 3 annual observations between 1981 and 2010 (a short record).

<table>
<thead>
<tr>
<th>Results</th>
<th>HBV</th>
<th>Ungauged catch.</th>
<th>Partially gauged catch.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>SVC</td>
<td>GS</td>
</tr>
<tr>
<td>RMSE (mm/yr)</td>
<td>394</td>
<td>315</td>
<td>389</td>
</tr>
<tr>
<td>ANE (1)</td>
<td>0.18</td>
<td>0.11</td>
<td>0.192</td>
</tr>
<tr>
<td>CRPS (mm/yr)</td>
<td>235</td>
<td>145</td>
<td>209</td>
</tr>
</tbody>
</table>

Main conclusions: (i) The SVC model performs better than the process-based model (HBV) and the geostatistical model (GS) for ungauged catchments. (ii) For partially gauged catchments a purely geostatistical method performs better, but the SVC model is indeed able to exploit short records.
References


