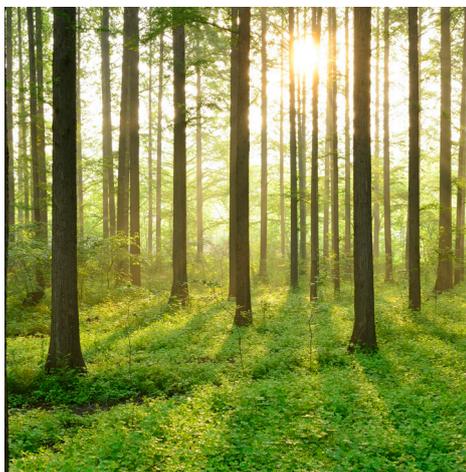


# SPRING FLOOD PREDICTIONS WITH HBV AND HYPE

REPORT 2021:777



HYDROLOGISKT  
UTVECKLINGSARBETE





# Spring Flood Predictions with HBV and HYPE

EDUARDO REYNOLDS, DAVID GUSTAFSSON & ILARIA CLEMENZI

ISBN 978-91-7673-777-4 | © Energiforsk May 2021

Energiforsk AB | Phone: 08-677 25 30 | E-mail: [kontakt@energiforsk.se](mailto:kontakt@energiforsk.se) | [www.energiforsk.se](http://www.energiforsk.se)



## Foreword

**Energiforsk's Hydrological development programme has long been interested in the topic of how to better implement the use of snow measurements to improve forecasts. An important question in this respect is whether a distributed model would perform better than a non- or semi-distributed model. The idea for this project was therefore developed, to compare the HYPE and HBV models.**

Behind this idea was the hypothesis that significant differences between the models might be found. However, as the reader will discover, there were no significant difference in performance between the two for the purposes of developing hydrological forecasts for the hydropower industry. This project instead gave valuable opportunity to explore how different types of snow measurements could be integrated to better describe spring flows. In the report the researchers have focused on the ability of the two different models to describe snow accumulation, snow melt, and spring flows.

The project, Spring Flood Determination with HBV and HYPE, has been carried out by Eduardo Reynolds together with David Gustafsson and Ilaria Clemenzi at SMHI.

The project has been part of HUVA - Energiforsk's working group for Hydrological development, the focus of which is to improve the precision of the hydropower industry's inflow forecasting models. From the HUVA working group a reference group for the project was formed consisting of Beatriz Quesada Montana and Peter Nordlander, UNIPER, Mikael Sundby, Vattenfall, Anna Hedström Ringvall and Björn Norell, Water Regulation Enterprises, Knut Sand, Statkraft, Johan Andersson, Fortum, Magnus Jämting, Jämtkraft, and Elina Mikaelsson, Holmen.

These are the results and conclusions of a project, which is part of a research programme run by Energiforsk. This work was also supported by Swedish Energy Agency under the research project 'Snow Distribution and Data Assimilation for improved spring flood forecast and Sustainable hydropower reservoir regulation' (46424-1). The author/authors are responsible for the content.

## Summary

Generally, the application of hydrological models for real-time operations relies on using discharge observations for calibration, verification, and correction of the model states before producing any runoff forecast. The main purpose of this study was to assess how different types of snow measurements can be integrated and used in hydrological modelling to better describe spring flows. The first part of the experiment was designed to explore the value of including snow measurements during the calibration procedure, and the second part of the experiment was to assess how long-term spring flow forecasts would improve if snow information was included during the forecast model initialization procedure (i.e. data assimilation). In this report we focus on a comparison of the HBV- and HYPE-model versions, currently in use by the Swedish Meteorological Institute, regarding their ability to describe snow accumulation, snow melt, and spring flows. The sub-basin of lake Överuman, located in the Ume river, was the area of study for this experiment and the snow data available was snow water equivalent (SWE) measurements and satellite products, such as fractional snow cover data (FSC).

The most important conclusions from this study were:

1. When snow data was included or excluded in calibration, the model performance was relatively similar for both the HBV and HYPE models with regards to their ability to describe snow storage and spring flows. However, including snow data in the calibration did not have any significant impact on the robustness of the model performance between calibration and verification periods.
2. The added value of the snow data was mainly seen in terms of improved forecast performance when snow data was assimilated during forecast model initializations.
3. Using reservoir inflow, snow water equivalent, and fractional snow cover data individually or in combination for data assimilation in HBV and HYPE generally improved the long-term spring flow forecast for all years and forecast periods from April-July.
4. The combinations of assimilating both inflow and snow water equivalent data provided the highest forecast performance.

## Keywords

HBV, HYPE, calibration, data-assimilation, long-term forecasts, local inflow, snow water equivalent, fractional snow cover

## Sammanfattning

Användningen av hydrologiska modeller för simulering och prognostisering av avrinning till sjöar och vattendrag förlitar sig vanligtvis på att endast använda observationer av avrinning (eller vattenföring) för kalibrering, verifiering och initialisering. Huvudsyftet med denna studie var att bedöma hur olika typer av snöobservationer kan användas för att förbättra modellering och prognostisering av avrinningen under vårfloden. I den första delen av experimentet undersöktes värdet av att inkludera snöobservationer i kalibreringen av hydrologiska modeller, och i den andra delen utvärderades om långtidsprognoser av avrinning kan förbättras genom att även inkludera snöobservationer för korrigering av prognosmodellernas starttillstånd (så kallad data-assimilering). I studien jämfördes de hydrologiska modellerna HBV- och HYPE med avseende på deras förmåga att beskriva snötäcke, snösmältning och vårflöden och förmåga att tillgodogöra sig informationen från snöobservationerna. Studieområde för undersökningen var tillrinningsområdet för sjön Överuman i övre delen av Umeälven, där det förutom tillrinningsobservationer även finns observationer av snövattenekvivalent genom årliga snötaxeringar med markradar, samt snötäckningsgrad för satellitbaserade observationer.

De viktigaste slutsatserna från studien listas nedan:

1. Modellernas förmåga att beskriva vårfloden efter kalibrering var relativt lika, men påverkades i detta experiment mycket lite av om snödata inkluderades eller inte i kalibreringen.
2. Mervärdet av snöobservationerna visade sig mest genom förbättrad prognoskvalitet när de inkluderades för korrigering (assimilering) vid initialisering av prognosmodellerna.
3. Assimilering av tillrinning, snövattenekvivalent och snötäckningsgrad individuellt eller i kombination under initialiseringen av HBV- och HYPE-prognosmodellerna förbättrade generellt vårflödelångtidsprognoserna för alla år och prognosperioder från april-juli.
4. Kombinerad assimilering av tillrinning och snövattenekvivalent gav de högsta förbättringen av prognoserna.

## Nyckelord

HBV, HYPE, kalibrering, data-assimilering, långtidsprognos, lokala flöden, snövattenekvivalent, snötäckningsgrad

## List of content

<b>1</b>	<b>Background</b>	<b>7</b>
<b>2</b>	<b>Material and Methods</b>	<b>8</b>
2.1	Study site	8
2.2	Hydrological Models	9
2.2.1	HBV	9
2.2.2	HYPE	13
2.2.3	Spatial set-up of HBV and HYPE models	16
<b>3</b>	<b>Experimental design</b>	<b>18</b>
3.1	Model calibration	18
3.2	Data assimilation and long-term SPRING FLOOD forecasts	21
3.2.1	HBV auto-correction tool	22
3.2.2	HYPE data assimilation with Ensemble Kalman Filter	22
3.2.3	Experiment on data assimilation and long-term forecasts	23
<b>4</b>	<b>Results</b>	<b>26</b>
4.1	Integration of snow information in calibration	26
4.2	Spring flood forecasting and data assimilation	35
4.2.1	Impact on spring flood forecasts when using snow in calibration	36
4.2.2	Impact on spring flood forecasts when using inflow and snow in data assimilation	38
<b>5</b>	<b>Discussion</b>	<b>41</b>
<b>6</b>	<b>Conclusions</b>	<b>45</b>
<b>7</b>	<b>Follow-up</b>	<b>46</b>
<b>8</b>	<b>References</b>	<b>47</b>
<b>Appendix A: Stage-volume table for Lake Överuman and SWE data at study site.</b>		<b>49</b>
<b>Appendix B: Additional results on model calibration when integrating snow data</b>		<b>50</b>
<b>Appendix C: Additional results on data assimilation and long-term spring flood forecasts</b>		<b>54</b>

# 1 Background

Snow cover within basins can vary widely and depend on height, wind exposure, land use, among other properties. Studies dealing with snow redistribution in hydrological models have shown that its consideration improves the simulation of snow patterns, snow water equivalent (SWE) and the prediction of discharge in mountain basins (Freudiger et al., 2017). Representative descriptions of distribution of snow within basins are expected to be achieved by using hydrological models with high spatial resolutions, by including snow information during calibration and/or during the updating procedure of the model states before producing seasonal and forecast products.

The latter updating procedure is referred here as data assimilation. Previous and ongoing Energiforsk projects (Gustafsson et al., 2015, Johansson et al., 2015, SMHI & UU, 2020) have shown the potential for improving spring-flood forecast if snow information could be included in data assimilation. Furthermore, the use of models with spatial resolutions similar to those from satellite products simplifies calibration and correction of model states against observations. For instance, Johansson et al. (2015) reported a model that describes redistribution of snow at a fine spatial resolution to describe winter precipitation in years with an unusual precipitation distribution. They applied a combination of the HBV and HOPE hydrological models to four basins in the Swedish mountains at a spatial resolution of 4X4 km<sup>2</sup>. Their results were equivalent to those of the HBV model used for operational spring-flood forecasts in Sweden, but the added spatial resolution and meteorological data did not lead to a clear improvement in model performance.

The main purpose of this project is to further explore how different types of snow measurements can be integrated and used in hydrological modelling to better describe spring flows. Our assumption is that including more than one variable in the calibration will improve the chance to get more a robust or consistent model when transferring parameters to a verification or forecasting period. In this report we focus on comparing the HBV- and HYPE-model versions, currently use by SMHI, regarding their ability to describe snow accumulation, snow melt, and spring flows.

The following report starts by comparing both models using local calibration against local inflow. Following, the added value for model performance and predictability after integrating snow measurements in the calibration and in different data assimilation procedures before running the models to produce long-term spring flow forecasts are explored. Here, snow information was available as snow water equivalent measurements and by means of satellite products, such as fractional snow cover data.

## 2 Material and Methods

### 2.1 STUDY SITE

The Ume River, or Umeälven, is one of the main rivers in northern Sweden. It is around 460 km long and has been extensively cultivated for hydropower production. The main channel of the Ume river is regulated and dammed from the larger spring lakes in the mountains and downstream to its outlet to the Gulf of Bothnia (Bottenviken). The sub-basin of lake Överuman is at the origin of the Ume River by the Norwegian border within the Scandinavian mountain range (SWEREF99 TM. N: 7314876. E: 500958) and was the study area for this experiment. It has a total area of 653 km<sup>2</sup> whereas the lake covers a surface area of 89 km<sup>2</sup>. The basin is main covered by heathland and it is a bare mountain region (Fig. 1). Geographical information, areas and the distribution between lake, open land and forest in different altitude zones were retrieved from SMHI's sub-basin area database SVAR.

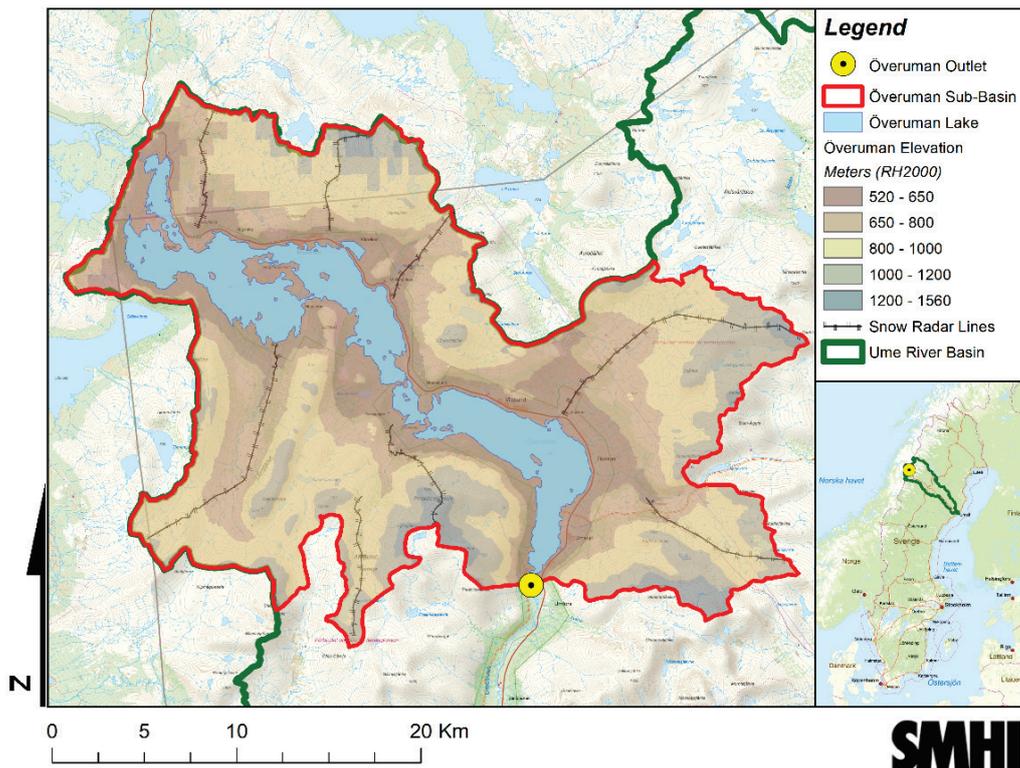


Figure 1. Location of the Överuman Sub-Basin.

Daily precipitation and temperature data are available with spatial resolution of about 4x4 km<sup>2</sup> (1981–2020) from SMHI's database based on interpolated temperature and precipitation observations using the PTHBV model (B. Johansson, 2000, B. Johansson & Chen, 2003). Hourly wind direction and air temperature data was available with spatial resolutions 22x22 km<sup>2</sup> (1998–2006), 11x11 km<sup>2</sup> (1998–

2006), and 2,5x2,5 km<sup>2</sup> (2016–2020) from SMHI's MESAN analysis based on the HIRLAM (1998–2015) and AROME (2015–2020) meteorological models and observations, respectively (Häggmark et al, 2003). Daily average values of observed outflow from, and water stages at, the Överuman lake were available for two stations in the basin from 1965 (i.e. stations 1856 and 1435 respectively). These two were used to compute the observed local inflow to the lake according to a water stage-volume table available for the Överuman lake (Table 1 in Appendix A). Here, the total inflow to the lake is equal to its local inflow since Överuman has no upstream basins. It is worth noting that the reliability of the discharge and water stage data in Överuman has been under discussion in recent years because an already established and calibrated model for Överuman began to overestimate after the Klippen's power station was put in operation in the mid-90s.

Additionally, different types of snow measurements such as fractional snow cover (FSC) and snow water equivalent (SWE) were available for this study. Estimates of SWE data were available for the years 2017–2020 (Table 2 in Appendix A). This data was derived by Vattenregleringsföretagen AB (VRF) from snow depth measurements within the Överuman basin taken along survey lines in intervals of about 1 m using ground penetrating radar. These estimates were spatially aggregated and used in HBV and HYPE for model calibration, verification, and data assimilation.

Fractional snow cover determined from the EU FP7 project CryoLand ([www.cryoland.eu](http://www.cryoland.eu); Schweizer et al, 2017) was further used for calibration and verification. This data is derived from optical satellites (mainly the North American MODIS) and comes with a spatial resolution of about 500x500 m<sup>2</sup>. Fractional snow cover data was downloaded (i.e. 2000–2020) and aggregated to the Överuman drainage area.

Data based on optical technologies such as the Cryoland is limited due to clouds and low sun angle, especially in Scandinavia during large parts of the winter. Because of this limitation, the FSC downloaded data was first filtered and quality controlled before used in this study. For example, FSC data was flagged and not used in the analysis for those days when more than 75% of the data was missing within the basin either because of clouds or temporary errors. Additionally, FSC values lower than 0,05 were flagged and not considered in the analysis to take away some weight during the summer periods when less information is added. Besides the latter, a visual quality control was carried out and uncertain FSC values based on temperature were flagged and not used in the analysis (e.g. high FSC values when they are expected to be low or vice versa, or reduction in FSC values when temperature has not increased and is still below 0° C).

## 2.2 HYDROLOGICAL MODELS

### 2.2.1 HBV

The HBV model is a conceptual hydrological model for continuous calculation of runoff. It was originally developed at the Swedish Meteorological and Hydrological Institute (SMHI) in the early 70's to assist hydropower operations by

providing hydrological forecasts (Bergström, 1976, Lindström et al., 1997). The first operational forecasts using HBV were carried out for basins in the northern part of Sweden in 1975 and nowadays the model is used as standard forecasting tool for the hydropower industry in Sweden. The HBV model has low data requirements and low computation demands. The HBV approach has proved flexibility and robustness to solve water-resource problems, and applications of the model have been reported in more than 80 countries with different hydrological and climatological conditions showing overall good results (Reynolds et al., 2017, 2020, Seibert, 1997, 1999).

The HBV has a user interface called Integrated Hydrological Modelling System (IHMS) and development of the model has taken place within Energiforsk projects in recent years (Johansson et al., 2015). It can be linked with Real Time Weather Information and Forecasting Systems, such as the HYFO and WebBIHMS systems developed by SMHI. Over the years only minor changes in the basic model structure have been made. In the beginning of the 1990s a comprehensive re-evaluation of the HBV model routines was carried out and resulted in the HBV-96 version (Lindström et al., 1997), which is the model version currently used within IHMS and for this study.

The model requires precipitation, air temperature, and potential evaporation as input to simulate river discharge at the basin scale. Normally, only temperature and precipitation are needed for Swedish basins since potential evaporation is calculated by the HBV model using a simplified variation of Thornthwaite's equation.

Furthermore, the model consists of subroutines for snow accumulation and snow melt, a soil moisture accounting procedure, routines for runoff generation and finally, a simple routing procedure. It is possible to run the model separately for several sub-basins and then the contributions from all sub-basins are added. Subdivisions into elevation zones can be made for basins of considerable elevation range as is the case for Överuman. This subdivision is only made for the snow and soil-moisture routines and each elevation zone is further divided into different vegetation zones (forested and non-forested areas). A schematic sketch of the HBV 96 model version is shown in Figure 2.

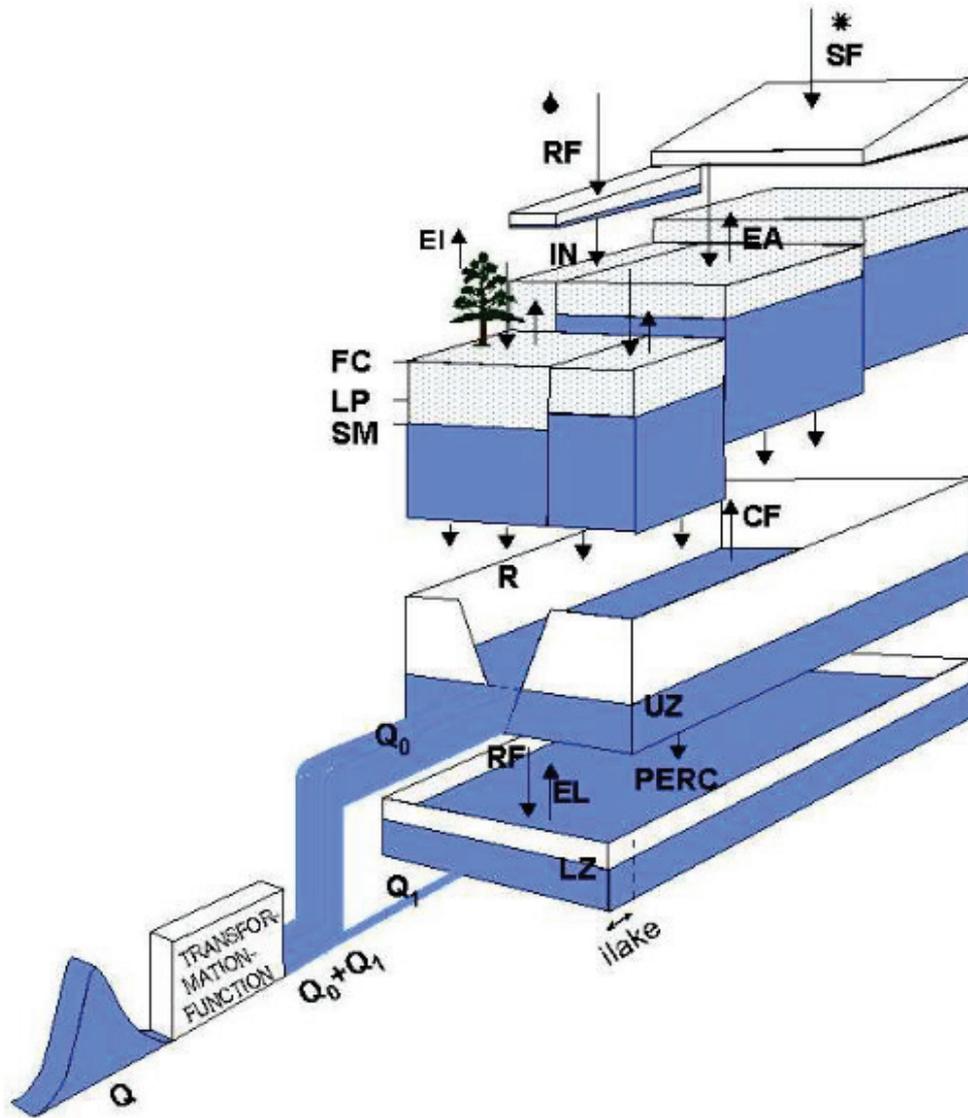


Figure 2. Schematic presentation of the HBV model for one sub-basin.

The HBV model has several parameters for tuning but here we limit the description of the model around those calibrated in this study (highlighted in **bold**):

1. First, all precipitation is multiplied by a general precipitation correction factor, **pcorr** [-], which represents systematic errors in the input data and missing evaporation in the model.
2. Then, precipitation is simulated to be either snow or rain depending on whether the temperature is above or below a threshold temperature, **tt** [°C].
3. Snowmelt is calculated based on the degree-day method using the **cfmax** melting factor [mm °C<sup>-1</sup>]. Meltwater and rainfall are retained within the snowpack until they exceed a certain fraction, **cwh** [-], of the water equivalent of the snow.
4. Liquid water within the snowpack refreezes according to a refreezing coefficient, **cfr**.
5. A snow fall distribution can be made in each elevation/vegetation zone by subdividing them in snow classes using the parameter **sclass**. This parameter determines the number of subareas with different snow accumulation. The distribution is given by the parameters **sfdistfo** and **sfdistfi**.
6. Rainfall and snowmelt are divided into water filling the soil box and groundwater recharge based on the current water content of the soil box (**SM** [mm]), the maximum soil-moisture storage (**fc** [mm]) and a shape factor (**beta** [-]).
7. Potential evaporation (PE) equals **athorn** [-] multiplied by the actual air temperature if the latter is larger than 0 °C, otherwise it equals 0 mm day<sup>-1</sup>.
8. Actual evaporation (EA) from the soil box equals PE if **SM/fc** is above **LP** [-] while a linear reduction of PE is used when **SM/fc** is below **LP**.
9. Groundwater recharge is added to the storage in the upper reservoir, **UZ** [mm] and as long as there is water in it, water will percolate to the lower reservoir, **LZ** [mm] according to the parameter **perc** [mm day<sup>-1</sup>].
10. In periods with high groundwater recharge, percolation is not sufficient to keep the upper reservoir empty and the generated discharge will have a direct contribution from the upper reservoir which represents drainage through more superficial channels. The lower reservoir, on the other hand, represents the groundwater storage contributing to the base flow.
11. Outflow from the upper reservoir is computed by a non-linear function defined by a continuously increasing recession coefficient, **k** [day<sup>-1</sup>] and the exponent **alpha** [-].
12. The parameters **khq** [day<sup>-1</sup>] and a reference discharge value, **hq** [mm day<sup>-1</sup>] are used in the model to calculate a value of **k** so that **hq** equals **khq** multiplied by **UZ<sub>hq</sub>**. **hq** is a high flow value at which the recession rate **khq** is assumed. In mountainous regions in Sweden such as in Överuman, **hq** is usually used as 7,0 mm day<sup>-1</sup>.
13. Outflow from the lower reservoir is computed as a linear function of the storage and the recession coefficient **k4** [day<sup>-1</sup>].
14. Finally, the total simulated runoff (the sum of the outflow from the upper and lower reservoirs) is routed on the following time steps by a triangular transformation function defined by the parameter **maxbaz** [day].

As previously stated, the Överuman basin has already an established model which has been calibrated against Q-data for a period before the Klippen's power station was put in operation and nowadays this calibrated model is used for producing short-term and long-term forecasts for the hydropower industry. Parameter values and model performance by this established and calibrated model for Överuman are used as reference benchmarks for comparison to the calibration of the HBV model performed here. It is worth to note that the calibration of the HBV model performed in this experiment is based on observations from the Klippen's power station.

### 2.2.2 HYPE

The HYPE model is a dynamic conceptual hydrological model that integrates water fluxes, nutrients, and other substances to describe processes in basins (Lindstrom et al., 2010). It is developed and maintained by SMHI. The model is coupled to HYSS which is a platform for managing inputs and outputs, optimization, and discretization in time and space. HYPE and HYSS are continuously developed, and new versions are released every two years. HYSS does not have a graphical user interface, which simplifies the possibilities for continuous development of the simulation system and the models coupled to it, but at the same time imposes greater demands to the users.

The model consists of several subroutines where the different hydrological processes are described. The landscape can be divided in different units (sub-basins) and each unit is further divided into sub-units which represent different soil and land use classes, linking the landscape characteristics to the physical processes by means of the model parameters. Snow accumulation and melt, soil processes, evaporation, and runoff are simulated individually for each land sub-unit in the sub-basins. The runoff from the land sub-units is further routed through surface water classes representing the river network and lakes towards the outlet of sub-basins.

Spatial distribution of snow within and between the sub-units of a sub-basin is represented by three different approaches in the HYPE model (described in more detail in the list below). First of all, the precipitation within a sub-basin is transformed into snowfall as a function of air temperature, which in turn is adjusted by elevation individually for each sub-unit. Sub-basins with large variability in elevation may thus get different amount of snowfall between the sub-units. Moreover, snow-melt rates are also affected by the elevation induced temperature differences during snow melt periods. Secondly, the snowfall may be re-distributed between the sub-units based on wind direction and topography. In addition to the elevation and wind induced snow distribution processes, the HYPE model also has an implicit representation of snow re-distribution using a so-called snow depletion curve. The snow depletion curve predicts how the fraction of snow-covered area within a sub-unit is changing during the winter as a function of snowfall and snowmelt.

A schematic description of the HYPE model is reported in Figure 3.

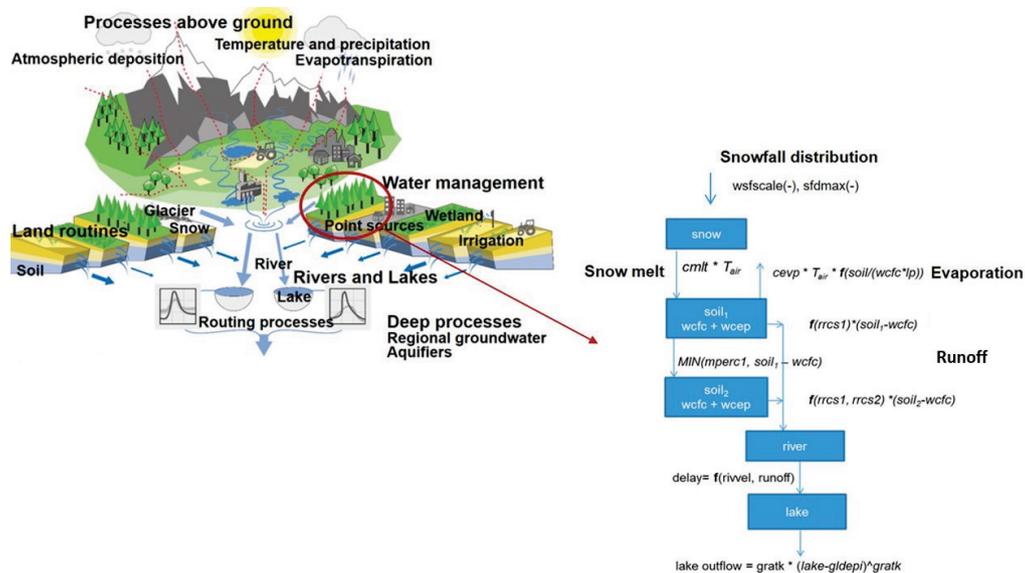


Figure 3. Schematic representation of the HYPE model.

In HYPE, hydrological processes are represented in a parametrized way and here we report the model parameters used in model calibration (highlighted in **bold**):

1. Similar to HBV, snow accumulation depends on whether precipitation falls as snow or rain. A temperature threshold, **ttpd** [°C] and a range **ttpi** [°C], defines the transition from 100% rain to 100% snow.
2. In the standard HYPE model setup, differences in the snow storage between the sub-units in a sub-basin are determined by elevation dependent air temperature differences, by vegetation-type dependent differences in a static snowfall correction factor and by the snow melt parameters. The purpose of the vegetation-type dependent snowfall correction is to represent impacts of interception on snow storage. No parameters related to these processes were included in the calibration procedure.
3. Here, we implemented an explicit distribution of snowfall based on the wind shelter factor concept (Winstral et al., 2002) between the sub-units in a sub-basin. This routine represents the snow spatial variability induced by the interaction between wind and topography occurring typically in mountainous environments. Theoretically, snowfall is distributed within the model sub-units by a wind shelter factor, WSF, which accounts for if a sub-unit is sheltered or exposed to wind, controlling the amount of snow falling on them. The wind shelter factor WSF is a physical characteristic input variable, derived separately for each sub-unit and each sub-basin from high resolution topography data. It represents the slope to the horizon in the wind direction and was derived for 8 wind directions using a 25x25 m<sup>2</sup> digital elevation model based on the ArcticDEM dataset (<https://www.pgc.umn.edu/data/arcticdem/>). The snowfall is distributed following the relation:

$$\frac{S_{sub-unit}}{S_{mean}} = w_N \cdot 10^{(wsf_{scale} \cdot WSF_{sub-unit})} \quad (1)$$

where  $S_{mean}$  is the mean snowfall in the model unit (sub-basin),  $S_{sub-unit}$  and  $WSF_{sub-unit}$  are the corrected snowfall and wind shelter factor for the sub-unit, respectively. The **wsf<sub>scale</sub>** [-] is a parameter related to the snowfall distribution induced by wind, topography and vegetation interaction. The **sfd<sub>max</sub>** [-] is a maximum value with which the snowfall distribution function will be truncated before the normalization and  $w_N$  is a set of wind shelter weights calculated for each sub-unit and sub-basin so that the negative and positive snowfall corrections add up to zero within each sub-basin.

4. Conceptually, the distribution of snowfall in HYPE based on the wind shelter factor concept is similar to the snow accumulation classes in the HBV model. The main difference is that in the HYPE model the snowfall distribution is based on current wind direction and physical characteristics derived from topography data, whereas in the HBV model this distribution is based on the empirical distribution of the parameters.
5. An implicit representation of redistribution of snow caused by wind is available by the calculation of a fractional snow cover within each sub-unit using the snow-depletion functions from Samuelsson et al (2006). The fractional snow cover is calculated as a function of the current snow storage in relation to the maximum snow accumulation during the current winter. The fractional snow cover within the sub-units is then used to scale the estimated snow melt and evaporation using the **fsceff** [-] parameter. The mean fractional snow cover in a HYPE model sub-basin is a summary of the fractional snow cover in the sub-units weighted by their areal fractions. If the HYPE model is run without the sub-unit fractional snow cover, then the sub-basin fractional snow cover will only depend on the occurrence of snow within the sub-units based on the explicit snowfall and snowmelt distribution as described in point 2-3.
6. A degree-day method is adopted to simulate snowmelt using the **cm<sub>lt</sub>** [mm °C<sup>-1</sup>] melting factor. The melting factor depends on land use and is calibrated for open land, shrub, and forest.
7. Melt water within the snowpack refreezes according to a refreezing coefficient, **cm<sub>refr</sub>**.
8. The infiltration and percolation of melt water and rainfall in soil are controlled by the parameters of wilting point, **wc<sub>wp</sub>** [-], field capacity, **wc<sub>c</sub>** [-] and effective porosity, **wc<sub>ep</sub>** [-].
9. Evaporation is calculated based on the modified Jensen-Haise/McGuinness model and is corrected by the parameter **ce<sub>vp<sub>corr</sub></sub>** [-].
10. The surface runoff occurring when the water table reaches the surface of the in the upper soil layer depends on the recession coefficient **sr<sub>rcs</sub>** [d<sup>-1</sup>].
11. Runoff from the topmost and bottom layer depends on the recession parameters **rr<sub>cs1</sub>** [d], **rr<sub>cs2</sub>** [d], respectively.
12. Within each sub-basin, the total runoff from the different soil layers in the sub-units is first routed through a local river using a delay function in days that is determined by the river length, and the water's maximum velocity, **riv<sub>vel</sub>** [m s<sup>-1</sup>] and a damping parameter, **damp** [-].

13. A fraction of the local river runoff may pass through a local lake, which stage and outflow are controlled by a standard rating curve. The main stream of the sub-basin collects outflow from the local stream/lake and from any upstream sub-basin, and the total inflow to the main stream is routed towards the sub-basin outlet with a similar function as the local stream.
14. An outlet lake may be present to control the sub-basin outflow using a standard rating curve or a more elaborate representation of reservoir regulations.
15. The HYPE model cannot directly provide local reservoir inflow corresponding to the observations as defined above in the study site section. Instead, local reservoir inflow is calculated by combining the calculated local runoff from the sub-basin covering the local drainage area of a reservoir together with the calculated net precipitation on lakes from the sub-basins covering the reservoir area, as described in Musuuza et al, (2020). Thus, delay in the main rivers and lakes in between connected reservoirs are disregarded, however this is of minor importance when analysing the total spring water runoff volumes.

### 2.2.3 Spatial set-up of HBV and HYPE models

In terms of the spatial representation of the landscape, hydrological models can be setup as lumped, semi-distributed and distributed. In a lumped model, spatial characteristics related to rainfall-runoff response over a basin are averaged and processes are represented as a single unit. In a semi-distributed model, areas in basins that respond hydrologically similar are divided into hydrologic response units (HRU) and processes are represented based on this division (i.e. the basin can be divided into different elevation and vegetation zones as well as into different sub-basins). In a distributed model, basins are divided on a particular grid size to capture spatial and temporal variability and then route flows through the basin from grid to grid. Explicitly, the latter type of models accounts for spatial variability using spatial datasets describing soils, vegetation, and land use.

For this study, both models were spatially set-up in a semi-distributed way. Additionally, we have developed and calibrated a spatially distributed HYPE model for the Överuman basin as part of the SNODDAS project. In that model, the Överuman basin has been divided in grids of 2.5x2.5 km<sup>2</sup>, corresponding approximately to the spatial resolution of the MESAN meteorological-analysis data which was used to obtain the input data on wind direction and air temperature. Furthermore, precipitation for the distributed HYPE model has been interpolated from the PTHBV data which has a spatial resolution of 4x4 km<sup>2</sup>. In both the semi-distributed and distributed HYPE model setups, the Överuman basin (unit) was further divided into 38 sub-units representing combinations of 3 elevation zones, 4 aspect zones (north, east, south, and west facing slopes), 5 land use classes (water, bare soil, shrubs, forest, and glacier) and 1 soil class.

As for the HBV, the model could also be spatially setup in a way where sub-basins are represented at finer resolutions or in a distributed way as it was done for HYPE, but this was not done in this experiment since Johansson et al. (2015) has already reported on the use of HBV in a distributed way and the use of the HBV model at a finer spatial resolution compared to the semi-distributed way currently

used for operational forecasting did not lead to a clear improvement in model performance.

This report is restricted to the analysis with the semi-distributed models, and comparisons between the semi-distributed and distributed HYPE models will come later as part of the final SNODDAS project outcomes.

### 3 Experimental design

The common practice for operational application of hydrological models is to use discharge observations for calibration, verification, and correction of the model states before producing any runoff forecast. The main purpose of this study was to assess how different types of snow measurements can be integrated and used in hydrological modelling to better describe spring flows. The first part of the experiment was designed to explore the value of including snow measurements during the calibration procedure whereas the second part of the experiment was to assess how long-term spring flood forecasts improves if snow information was previously included during the data assimilation procedure.

#### 3.1 MODEL CALIBRATION

For the first part of the experiment, the spatially semi-distributed models of HBV and HYPE were calibrated two times with assumptions of different types of information available for calibration:

- calibrating against only local inflow as typically done in practice (i.e. here the ability to describe water regulation is not evaluated),
- calibrating against local inflow and snow water equivalent information.

In addition to the previous, two additional calibrations were carried out with the HBV model:

- calibrating against local inflow and fractional snow cover data,
- calibrating against local inflow, snow water equivalent and fractional snow cover data.

For calibration, both models were run with 100,000 parameter sets (i.e. Monte Carlo simulations) and the most representatives, referred here as behavioral parameter sets, were selected for each calibration procedure based on model performance. The parameter sets used in calibration were randomly generated assuming a uniform distribution with predefined parameter-value ranges used in previous HBV and HYPE applications in many Swedish basins (Tables 1 and 2, respectively). Parameters for the spatially semi-distributed models were calibrated for the period 01 Oct 2008 – 30 Sep 2018, whereas 01 Oct 1998 – 30 Sep 2008 and 01 Oct 2018 – 31 Jul 2020 were used as verification periods.

Table 1. Parameter ranges used for calibration of HBV

Parameter	Description	Min–Max	Current parameter set	Unit
<b>Snow Routine</b>				
pcorr	General precipitation correction factor	0,8 – 1,2	1,3	-
tt	Threshold temperature	-2,0 – 2,0	-0,5	°C
cfmax	Melting factor	2,0 – 5,0	2,57	mm °C <sup>-1</sup>
<b>Soil Moisture Routine</b>				
fc	Maximum field capacity	50 – 150	100	mm
beta	Exponent in formula for drainage from soil.	1,0 – 4,0	1,67	-
athorn	Potential evaporation parameter	0,15 – 0,3	0,17	-
<b>Response Routine</b>				
perc	Percolation capacity from upper to lower response box	0,01 – 5,0	1,71	mm day <sup>-1</sup>
alpha	Non-linearity coefficient	0,0 – 1,0	1,0	-
khq	Recession coefficient for the upper box when water discharge equals hq.	0,005 – 0,5	0,39	day <sup>-1</sup>
hq	Reference discharge value	7,0	10,31	mm day <sup>-1</sup>
k4	Recession coefficient for lower response box	0,001 – 0,1	0,004	day <sup>-1</sup>
<b>Transformation Routine</b>				
maxbaz	Length of triangular transformation function	0,0 – 7,0		day

Table 2. Parameter ranges used for model calibration of HYPE.

Parameter	Description	Min–Max	Unit
<b>Atmosphere</b>			
ttpd	Snowfall/rainfall threshold	-2,0 – 4,0	°C
cevpcorr	Correction factor for evaporation	-1,0 – 1,0	-
wfscale	Scaling factor for snowfall distribution	0,01 – 0,1	-
sfdmax	Maximum amount of snowfall distribution	5,0 – 10,0	-
<b>Land routine</b>			
cmlt (open land, shrub, forest)	Snow melt factor	0,01 – 0,1	°C mm day <sup>-1</sup>
wcep	Effective porosity	0,01 – 0,2	-
srrcs (open, shrub, forest)	Recession coefficient for surface runoff	0,001 – 0,5	day
rrcs1	Recession coefficient for uppermost soil layer	0,2 – 0,7	day <sup>-1</sup>
rrcs2	Recession coefficient for lowest soil layer	0,03 – 0,2	day <sup>-1</sup>

Four objective functions were used to assess model performance regarding their ability to describe snow and spring flows:

1. Nash-Sutcliffe efficiency ( $R_2$ ),
2. volume-error of local inflow ( $VE_Q$ ),
3. volume error of SWE ( $VE_{SWE}$ ),
4. the Pearson-correlation coefficient of FSC ( $PearCorr_{FSC}$ ).

The first is a statistic measure commonly used to assess the goodness-of-fit of the simulated hydrographs, which tends to depend mostly on fitting periods with high flow conditions, the second and the third are indicators of the agreement between the averages of the simulated and observed local inflow (i.e. long-term water balance) and SWE respectively, whereas the fourth is a measure of the strength of the relationship between the observed and simulated FSC for the same period.

The values of the four objective functions were transformed into membership functions ( $X_1$ ,  $X_2$ ,  $X_3$  and  $X_4$ ) and then joined into four different joint functions ( $F1_Q$ ,  $F2_{Q,swe}$ ,  $F3_{Q,FSC}$  and  $F4_{Q,SWE,FSC}$ ). The 100 best parameter sets, with respect to each joint function were retained and considered as behavioural or representative to the Överuman basin. The distribution of the parameter values, as well as model performance and simulations achieved by the parameter sets selected for each

function were compared to assess the value added after integrating snow measurements in the calibration.

Equations for the different membership and joint functions are listed as follow:

$$X_1 = \begin{cases} R_2, & \text{if } R_2 \geq 0 \\ -999,99, & \text{otherwise} \end{cases} \quad (2)$$

$$X_2 = \begin{cases} VE_Q, & \text{if } VE_Q \leq 0,10 \\ -999,99, & \text{otherwise} \end{cases} \quad (3)$$

$$X_3 = \begin{cases} VE_{SWE}, & \text{if } VE_{SWE} \leq 0,30 \\ -999,99, & \text{otherwise} \end{cases} \quad (4)$$

$$X_4 = \begin{cases} 1 - PearCorr_{FSC}, & \text{if } PearCorr_{FSC} \geq 0,50 \\ -999,99, & \text{otherwise} \end{cases} \quad (5)$$

$$F1_Q = (W_{R_2} \times X_1) - (W_{VE} \times X_2) \quad (6)$$

$$F2_{Q,SWE} = (W_{R_2} \times X_1) - (W_{VE} \times X_2) - (W_{SWE} \times X_3) \quad (7)$$

$$F3_{Q,FSC} = (W_{R_2} \times X_1) - (W_{VE} \times X_2) - (W_{FSC} \times X_4) \quad (8)$$

$$F4_{Q,SWE,FSC} = (W_{R_2} \times X_1) - (W_{VE} \times X_2) - (W_{SWE} \times X_3) - (W_{FSC} \times X_4) \quad (9)$$

Where:

- $W_{R_2}$  is the weighting factor for  $R_2$  (here set as 1,00).
- $W_{VE}$  is the weighting factor for  $VE_Q$  (here set as 0,1).
- $W_{SWE}$  is the weighting factor for  $VE_{SWE}$  (here set as 0,3).
- $W_{FSC}$  is the weighting factor for  $PearCorr_{FSC}$  (here set as 0,3).

Typically values for  $W_{R_2}$  and  $W_{VE}$  used in HBV applications within SMHI were chosen for this experiment. For simplification, values for  $W_{SWE}$  and  $W_{FSC}$  were chosen to be the same but they could be different. A value equal to 0,3 was chosen for  $W_{SWE}$  and  $W_{FSC}$  because an earlier sensitivity analysis for different values of them (i.e. 0,1, 0,2 and 0,3) showed the largest impact in the selection of the parameter sets when this value was used for integrating snow in calibration.

### 3.2 DATA ASSIMILATION AND LONG-TERM SPRING FLOOD FORECASTS

For the second part of the experiment, we made use of the optimal parameter sets obtained for both models from each calibration procedure and then used different data assimilation approaches based on data availability before running long-term spring flood forecasts for the years 2017–2020. It is worth noting that four years may be a short period for a solid statistical analysis of the impact of the snow observations on forecast initialization and forecast skill.

Model states in the HBV can be updated or corrected against local inflow automatically (i.e. direct auto-correction method) or (and) manually based on adjustments to precipitation and temperature. When corrections to model states or variables were allowed before generating the long-term forecasts, corrections against local inflow were made automatically for the HBV, supplemented by manual correction of SWE. For the same procedure, HYPE used the Ensemble Kalman Filter (EnKF) which makes it possible to assimilate different types of observations simultaneously with the same method. Contrary to the auto-

correction method, model states which are not commensurable to the observations can also be updated with an EnKF approach based on several different types of observations (e.g. snow depth and fractional snow data).

### 3.2.1 HBV auto-correction tool

The auto-correction tool in HBV is based on the Brent's numerical method. The algorithm runs the model and searches for optimization windows where the absolute difference between measured and calculated flow is greater than an optimization parameter called mindiff (here, set as 0,1 in the model). If the difference is greater than this parameter, corrections are made to precipitation and temperature data until this condition is fulfilled otherwise not. After corrections are made for an optimization window, model continues searching for other optimization windows until it reaches the last day specified for auto-correction. Finally, the model is run for the whole period with the new corrections.

### 3.2.2 HYPE data assimilation with Ensemble Kalman Filter

The EnKF method can be used to assimilate any observation that can be predicted by the HYPE model, even observations that are not part of the model state variables. Essentially, the EnKF corrects the model state variables based on the covariance between errors in model state variables and errors in the model predictions of the assimilated observations.

The error co-variances are estimated by sampling from an ensemble of potential model states, predictions, and observations, generated by adding random perturbations to the model input variables and observations. The magnitude of the correction depends on the balance between the variance in the model state errors (governed by the assumed uncertainty in the meteorological forcing data) and the observational errors (governed by the assumed uncertainty in the assimilation data). A small error variance in the observation in relation to the variance of the model error, will result in a larger adjustment of the model towards the observations, and vice versa. However, if the uncertainty in the meteorological forcing variables is set too high, the data assimilation procedure may reduce the predictive ability of the model resulting in larger errors than a deterministic simulation (i.e. model runs without correcting model states or variables).

In this experiment, model state errors were generated by assuming relative standard deviations of 30% for precipitation error, 15% for wind speed error, and a constant standard deviation for air temperature error of 1°C. For observation errors, a relative standard deviation of 1% was assumed for SWE, while 2.5% was assumed for FSC, and a constant standard deviation of 5 m3s-1 was assumed for the local inflow. The resulting observational errors are probably too small compared to the real observational uncertainty.

### 3.2.3 Experiment on data assimilation and long-term forecasts

Four data-assimilation approaches were carried out with both models before generating the spring flood forecasts:

1. deterministic,
2. against local inflow,
3. against SWE,
4. against local inflow and SWE.

In addition to the previous, we carried out additional data-assimilation approaches with the HYPE model listed as follow:

1. Open (model runs using the same ensemble perturbations as in EnKF but without correcting model states or variables),
2. against FSC,
3. against local inflow and FSC,
4. against local inflow, FSC and SWE,
5. against SWE and FSC.

As previously described in section 3.1, the HBV model was calibrated with 4 different combinations of data: inflow, SWE and FSC data (i.e. joint functions F1–F4), whereas the HYPE model was calibrated only with 2 objective functions based on inflow and SWE data (i.e. F1–F2). Fractional snow cover was not used in calibration with the HYPE model as it was done with HBV model; however, the HYPE model also has a within-snow-class parameterization of fractional snow cover (refer to point 5 in section 2.2.2), which had been calibrated with the same data in a previous study (Musuza et al, 2020). The forecast experiments with HYPE(F1) and HYPE(F2) were thus repeated with and without this fractional snow cover function, and referred to as HYPE (Fx, off) and HYPE (Fx, on), respectively.

The different combinations of model calibration and data-assimilations for each model are summarized in Table 3. Note that for HBV, the optimal parameter set found for F2 and F4 was the same, therefore these models are referred to as HBV (F2,F4) since they provide identical results. Results for the parameter set currently used for the HBV operational are referred as Curr.

**Table 3. Combinations of models, calibration and data-assimilation approaches for forecast initialization (different combinations of inflow(Q), snow water equivalent (SWE) and fractional snow cover (FSC) for calibration and data assimilation during forecast model initialization). Curr stands for the current or established operational model at Överuman.**

Variables used in calibration	Label	Data-assimilation approach for forecast initialization								
		Det	Open	Q	SWE	Q SWE	FSC	Q FSC	Q SWE FSC	SWE FSC
<b>HBV</b>										
Q (Curr)	Curr	X		X	X	X				
Q	F1	X		X	X	X				

Variables used in calibration	Label	Data-assimilation approach for forecast initialization								
		Det	Open	Q	SWE	Q SWE	FSC	Q FSC	Q SWE FSC	SWE FSC
Q, SWE										
Q, SWE, FSC	F2,F4	X		X	X	X				
Q, FSC	F3	X		X	X	X				
<b>HYPE</b>										
Q	F1,off	X	X	X	X	X	X	X	X	X
Q, SWE	F2,off	X	X	X	X	X	X	X	X	X
Q, SWE, FSC	F2,on	X	X	X	X	X	X	X	X	X
Q, FSC	F1,on	X	X	X	X	X	X	X	X	X

Historical input data from 1998 until the previous year of the forecasted year were used to drive the model for the long-term spring flood forecasts. Data assimilation was allowed from 1st December of the previous year until the date the long-term forecasts were initiated. For each forecasted year, seven long-term forecasts were produced starting from the first of each month (i.e. 1st Jan, 1 Feb, 1 March and so on) until the 31st of July of such year.

The accumulated local inflow obtained for every year were compared to the observed local inflow for the same period to assess how the long-term forecasts improved by the different calibration and data-assimilation strategies. Long-term spring flood forecast results are presented in day equivalent units (DE= 86,400 m<sup>3</sup>). Furthermore, a climatological reference forecast, calculated as the mean observed local inflow for the seven forecast periods based on all previous historical years since 1965 was used to assess the added value of the different data-assimilation methods. The predictive ability of the long-term spring flood forecasts was based on relative volume error (RE) and the Continuous Ranked Probability Skill Score (CRPS, Hersbach, 2000). CRPS assess the ensemble forecast performance by comparing the cumulative distribution function of the forecast ( $P$ ) with a step function ( $H$ ) corresponding to the observations:

$$CRPS = \frac{1}{N} \times \sum_{i=0}^N (P_i(x) - H(x))^2 \quad (10)$$

where  $N$  is the number of observations i.  $H(x)=0$  when the predicted values are smaller than the observed ( $x$ ), and  $H(x)=1$  when the predicted values are larger than  $x$ . CPRS can be explained as the mean square error of the predicted cumulative distribution and is a probabilistic generalization of the mean absolute error.

The normalized version of the CRPS (i.e. CRPSS) was computed to compare the forecast system to a reference benchmark (in our case the climatological forecast):

$$CRPSS = 1 - \frac{CRPS_{Model\ Forecast}}{CRPS_{Benchmark}} \quad (11)$$

CRPSS is based on the square error of each ensemble member. CRPSS>0 means that the models are better than the climatological forecasts, and vice versa. CRPSS<0 means that the spread of the model ensemble is lower than the spread of the reference model.

## 4 Results

### 4.1 INTEGRATION OF SNOW INFORMATION IN CALIBRATION

Different calibration approaches based on the availability of different types of snow data were tested in this experiment. In terms of how well the parameters could be identified for HBV, all calibration approaches showed the same behavior (Figure 4). The distribution of the parameter values for  $p_{corr}$ ,  $cf_{max}$ ,  $tt$ ,  $khq$  and  $max_{baz}$  were relatively constrained and well identified, whereas the distribution for the other parameters was more spread. Here the parameter ranges used in calibration were kept in the y-axis. If the parameter value of the parameter set currently used in the operational model was outside this range, then the y-axis was further extended.

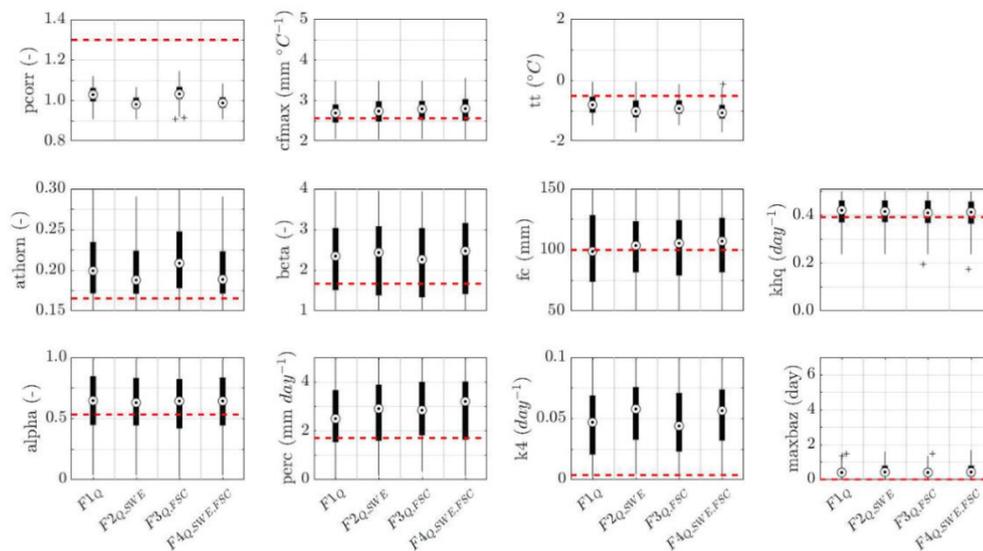


Figure 4. Boxplots of parameter values for the different calibration approaches of the HBV model.

Despite integrating different types of snow data in calibration (i.e.  $F2_{Q,SWE}$ ,  $F3_{Q,FSC}$  and  $F4_{Q,SWE,FSC}$ ), the distribution of the behavioural parameters was very similar to those calibrated only against inflow (i.e.  $F1_Q$ ). This was true even when  $F1_Q$  had more than half of the behavioural parameter sets different than those selected for  $F2_{Q,SWE}$  and  $F4_{Q,SWE,FSC}$  (i.e. more than 50). The calibrations  $F1_Q$ , and  $F3_{Q,FSC}$  had 75 parameter sets in common whereas  $F2_{Q,SWE}$  and  $F4_{Q,SWE,FSC}$  had 86 (Table 1 in Appendix B).

The parameter value for  $p_{corr}$  (i.e. the general precipitation correction factor) currently used in the operational model differs greatly with the median value obtained for the different calibration approaches carried out in this experiment (i.e. a value of 1,3 against a median value close to 1,0). This means that the established model assumes that the observed precipitation is 30% greater whereas the calibration carried out in this study assumes the observed precipitation to be relatively correct. The implications of the latter in terms of model performance and

predictability are later discussed. Other parameters such as  $\theta$ ,  $\beta$ ,  $\alpha$  and  $k_4$  also differed greatly between the established model and the different calibration approaches tested here but the distribution of those parameters in the latter was considerably spread and it would not be a fair comparison.

The distribution of the parameter values of the HYPE model showed that model parameters could be identified within the initially assigned range (Figure 5). Most of the parameters controlling the amount and distribution of snow in the basin (ttpd, wsfscale and cmelt for the different land use) and evaporation (cevpcorr) were rather well constrained compared to the parameters controlling runoff from the soil and surface water compartments. However, the effective porosity (wcep) and to some extent the recession coefficients for soil and surface runoff (i.e. srrcs and rrcs1) were also well constrained by the calibration data. Comparing the value distributions for the different criteria, even though small, some differences were found. The median of the temperature threshold, ttpd and the factor for snowfall distribution, wsfscale were lower when SWE was included in calibration than when calibrating only against inflow data.

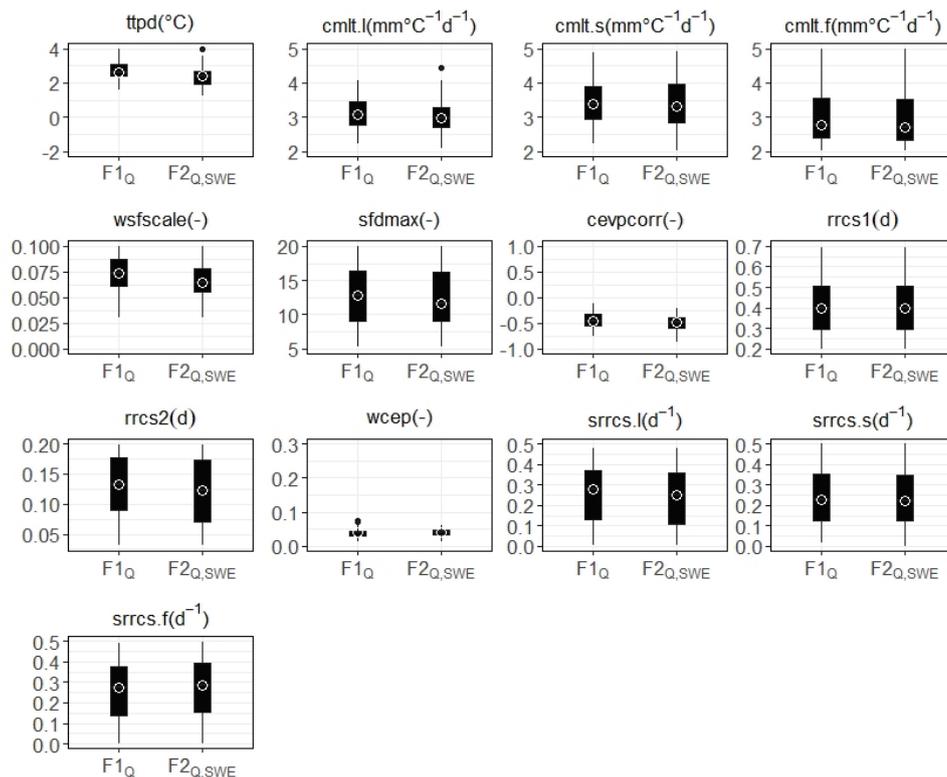
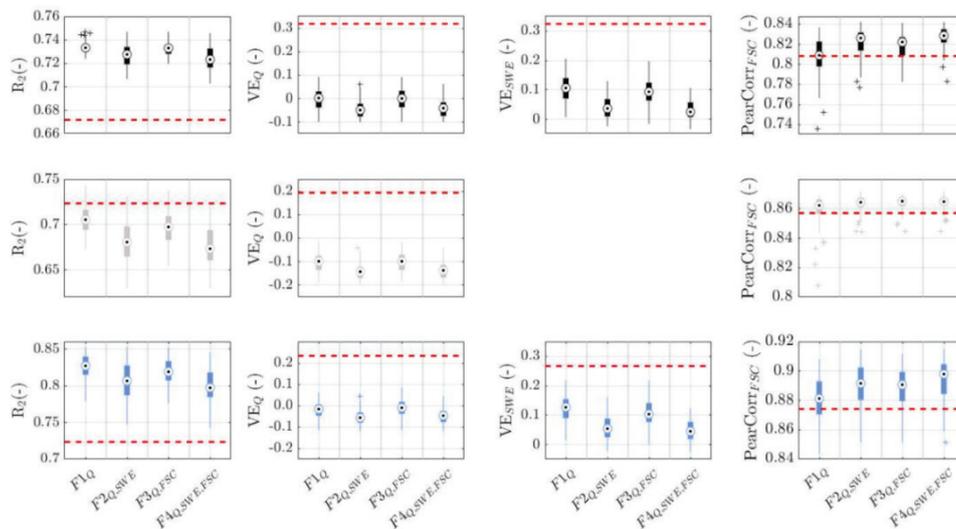


Figure 5. Boxplots of parameter values for the different calibration approaches of the HYPE model. The letters l, s and f in the snow melting factor (cmelt) and recession coefficient for surface runoff (srrcs) refer to the land use classes of open land, shrub and forest.

Model performance in the HBV model was somewhat similar for the different calibration approaches tested when snow data was or was not included in calibration (Figure 6). This was true for both calibration and verification. This doesn't come as a surprise considering that the distribution of parameter values presented in Figure 4 for HBV did not show much variability between the different calibration approaches.

Regardless of the period, model performance with respect to Nash-Sutcliffe efficiencies ( $R_2$ ) was typically better for  $F1_Q$ , volume error of the inflow data ( $VE_Q$ ) was better for  $F3_{Q,FSC}$ , whereas correlations coefficients for FSC ( $PearCorr_{FSC}$ ) were better for  $F4_{Q,SWE,FSC}$ . The best performances with respect to volume error of SWE ( $VE_{SWE}$ ) were found when SWE was integrated in calibration (i.e.  $F2_{Q,SWE}$  and  $F4_{Q,SWE,FSC}$ ) but at the cost of a slight decrease in performance of  $R_2$  and  $VE_Q$  from when the model was only calibrated against inflow (i.e.  $F1_Q$ ). For instance, in the calibration period  $VE_Q$  decreased from 0% when calibrating only against inflow to -5% when SWE was included in calibration whereas the opposite happened for  $VE_{SWE}$  (it improved from 11% to 4%). Similar patterns were seen for  $VE_Q$  and  $VE_{SWE}$  in the two verification periods. The spread in model performance in terms of  $PearCorr_{FSC}$  during all the periods was generally larger and not as good for  $F1_Q$  compared to the other calibration procedures that included some type of snow data in calibration. Pearson correlation coefficients of FSC were typically high (above 0.75).



**Figure 6. Model performance comparison for the different calibration procedures of the HBV model in calibration (20081001–20180930, first row) and in the verification periods (19981001–20080930 and 20181001–20200731, second and third row respectively).**

Model performances in calibration and in verification were typically better for the different calibration approaches tested in this experiment than to the HBV parameter set currently used in the operational model (except for  $R_2$  in the first verification period: 19981001–20080930). Generally, the parameter set used in the operational model overestimated considerably runoff volume which resulted in model performances with large errors with respect of  $VE_Q$  and  $VE_{SWE}$ . The latter was mostly caused by the large increase of the observed precipitation assumed in the established model (i.e. 30%). At the same time, this should not come as a surprise since it is already known that this model largely overestimates the observations after the Klippen’s power station was put in operation.

With respect to HYPE, the distribution of the model performance in relation to inflow ( $R_2$ ) was found ranging around 0,75 for the calibration period when the model was calibrated only against inflow  $F1_Q$ , and in combination with SWE  $F2_{Q,SWE}$  (Figure 7). The inflow volume error was found below 5% both for  $F1_Q$ , and  $F2_{Q,SWE}$ , and volume error of SWE was found to be smaller for  $F2_{Q,SWE}$  than for  $F1_Q$  during calibration.

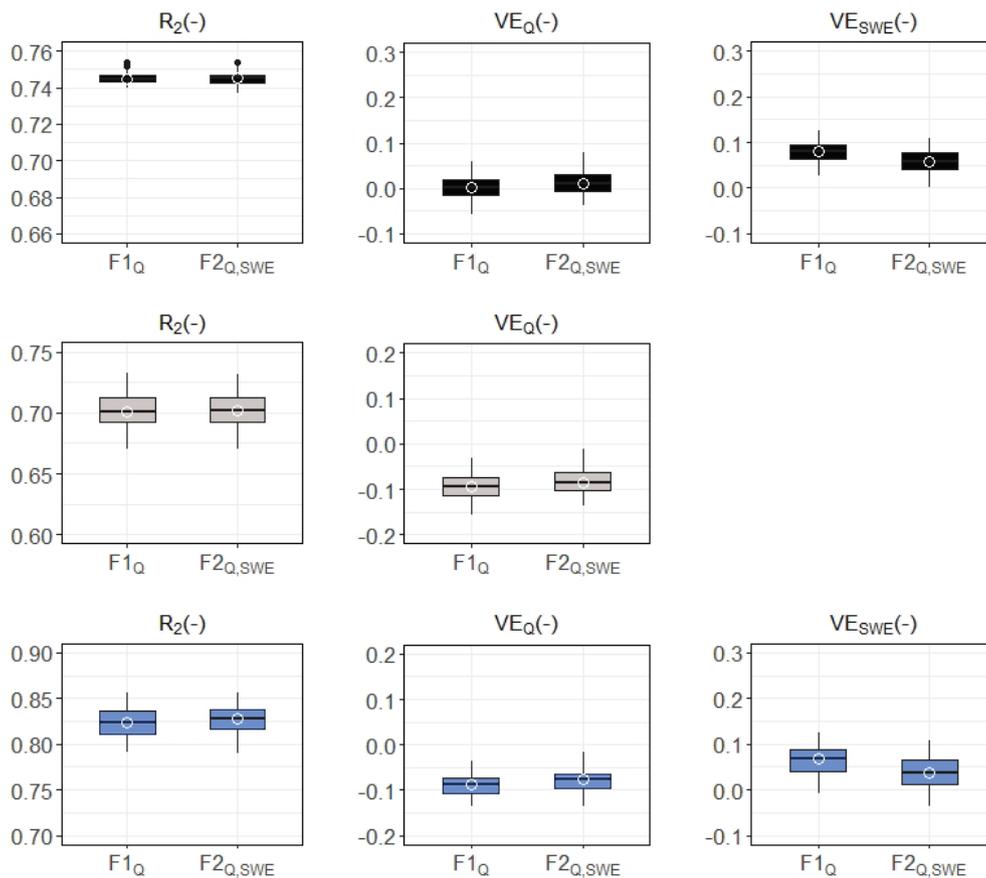


Figure 7. Model performance comparison for the different calibration procedures of the HYPE model in calibration (20081001–20180930, first row) and in the verification periods (19981001–20080930 and 20181001–20200731, second and third row respectively).

In the first verification period (1998–2008) model performance decreased in relation to inflow for both criteria (i.e.  $R_2$  and  $VE_Q$ ) compared to the calibration period. In the second verification period the model performance in relation to  $R_2$  was found higher compared to the first verification period. The inflow volume error was not as good in the verification periods compared to the calibration period with model underestimations between 5% and 10% for both calibration criteria. However, the distribution of the inflow volume error and SWE for  $F2_{Q,SWE}$  was shifted towards smaller errors indicating that including SWE observations had an effect on model calibration even though small.

A summary in terms of model performance by both models achieved in calibration (Table 4) and in the verification periods (Table 5 and Table 6) is shown next.

**Table 4. Median values of model performances for HBV and HYPE, spatially setup in a semi-distributed way during the calibration period (20081001–20180930).**

Calibration approach	$R_2$ (-)	$VE_Q$ (%)	$VE_{SWE}$ (%)	$PearCorr_{FSC}$ (%)
<b>HBV</b>				
$F1_Q$	0,73	0	11	0,81
$F2_{Q,SWE}$	0,73	-5	4	0,83
$F3_{Q,FSC}$	0,73	0	9	0,82
$F4_{Q,SWE,FSC}$	0,72	-4	3	0,83
<b>HYPE</b>				
$F1_Q$	0,74	0	8	
$F2_{Q,SWE}$	0,74	1	6	

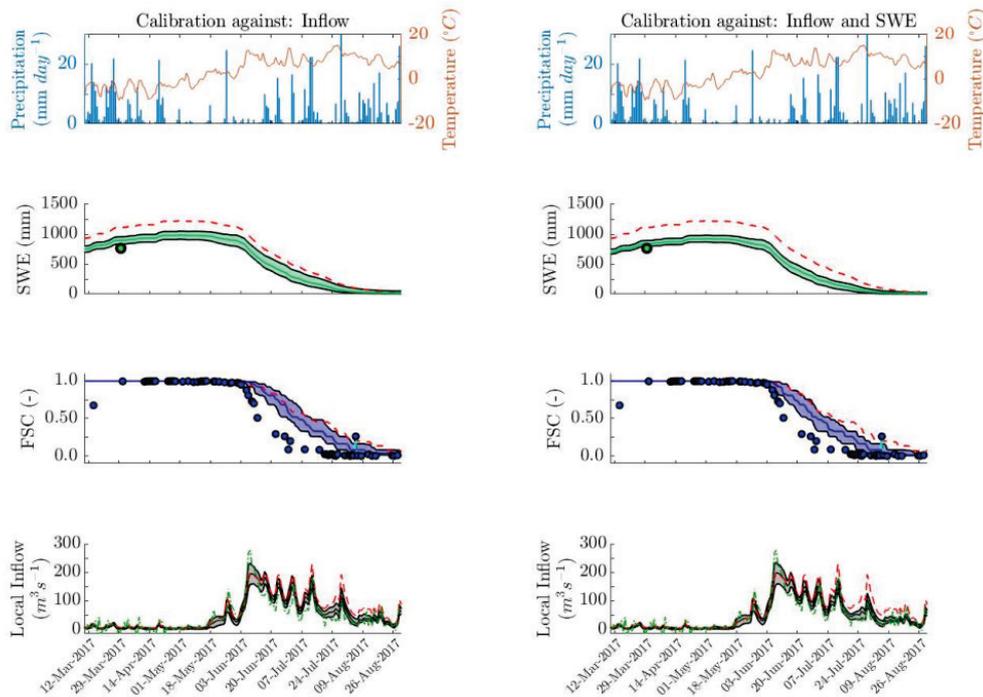
**Table 5. Median values of model performances for HBV and HYPE, spatially setup in a semi-distributed way during the first verification period (19981001–20080930).**

Calibration approach	$R_2$ (-)	$VE_Q$ (%)	$VE_{SWE}$ (%)	$PearCorr_{FSC}$ (%)
<b>HBV</b>				
$F1_Q$	0,71	-10	-	0,86
$F2_{Q,SWE}$	0,68	-14	-	0,86
$F3_{Q,FSC}$	0,70	-10	-	0,87
$F4_{Q,SWE,FSC}$	0,67	-14	-	0,86
<b>HYPE</b>				
$F1_Q$	0,70	-9	-	
$F2_{Q,SWE}$	0,70	-9	-	

**Table 6. Median values of model performances for HBV and HYPE, spatially setup in a semi-distributed way during the second verification period (20181001–20200731).**

Calibration approach	$R_2$ (-)	$VE_Q$ (%)	$VE_{SWE}$ (%)	$PearCorr_{FSC}$ (%)
<b>HBV</b>				
$F1_Q$	0,83	-2	13	0,88
$F2_{Q,SWE}$	0,81	-6	5	0,89
$F3_{Q,FSC}$	0,82	-1	10	0,89
$F4_{Q,SWE,FSC}$	0,80	-5	5	0,90
<b>HYPE</b>				
$F1_Q$	0,82	-9	7	
$F2_{Q,SWE}$	0,83	-8	4	

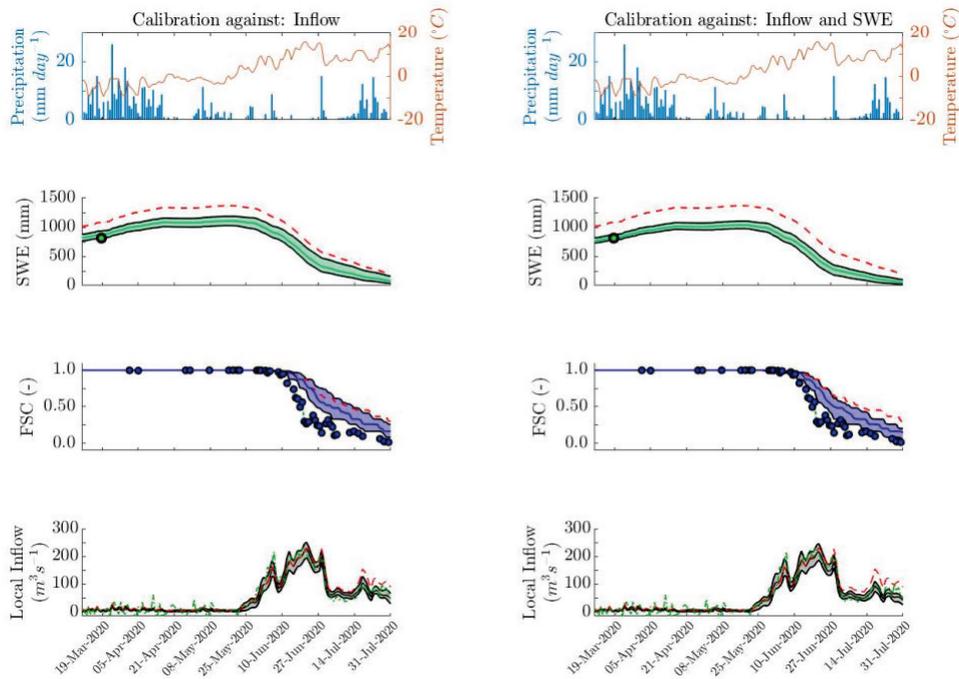
Our findings for HBV and HYPE when integrating snow information in calibration can be better illustrated by comparing simulations by the different calibration approaches (Figures 8 and 9 for HBV, whereas Figures 10 and 11 for HYPE, both during calibration and verification respectively. See also Appendix B).



**Figure 8.** HBV simulations by parameters calibrated against inflow,  $F1_Q$  and against inflow and SWE,  $F2_{Q,SWE}$  for the year 2017 during the calibration period. Green circles in the SWE subplots represent the SWE radar observations, blue circles in the FSC subplots represent the FSC observations, whereas the green dotted line in the local inflow subplots represent the observed local inflow. For all subplots the read dotted lines represent the simulations by current operational model. Local inflow, SWE and FSC simulations within the 5th and 95th percentile uncertainty bounds together with their median are illustrated within their respective subplots.

Generally, uncertainty bounds were slightly larger for those parameters calibrated against inflow than those calibrated against inflow and SWE but simulation results were relatively similar in terms of predictability (very narrow differences can be found between the two). Snow water equivalent seems to be slightly better simulated when snow information was included in calibration. Both models tended to overestimate the SWE observations in the year 2017 during calibration and, although with a minor effect, in the year 2020 during the verification period. This is reflected in the timing of the disappearance of the snow as illustrated by the fractional snow cover. In most cases, both models failed to reproduce the rapid decrease in the fractional snow cover towards the end of the snow melt period.

For the years 2018 and 2019, a good agreement between observations and model simulations was found (Appendix B). The calibration results shown for HYPE in this section are for when the fractional snow cover function was disabled (i.e.  $F_x$ , off). Simulations results by the established HBV operational model in Överuman showed large overestimations of SWE and local inflow as would it be expected from the model performance results previously shown.



**Figure 9.** HBV simulations by parameters calibrated against,  $F1_Q$  and against inflow and SWE,  $F2_{Q,SWE}$  for the year 2020 during the verification period. Green circles in the SWE subplots represent the SWE radar observations, blue circles in the FSC subplots represent the FSC observations, whereas the green dotted line in the local inflow subplots represent the observed local inflow. For all subplots the read dotted lines represent the simulations by current operational model. Local inflow, SWE and FSC simulations within the 5th and 95th percentile uncertainty bounds together with their median are illustrated within their respective subplots.

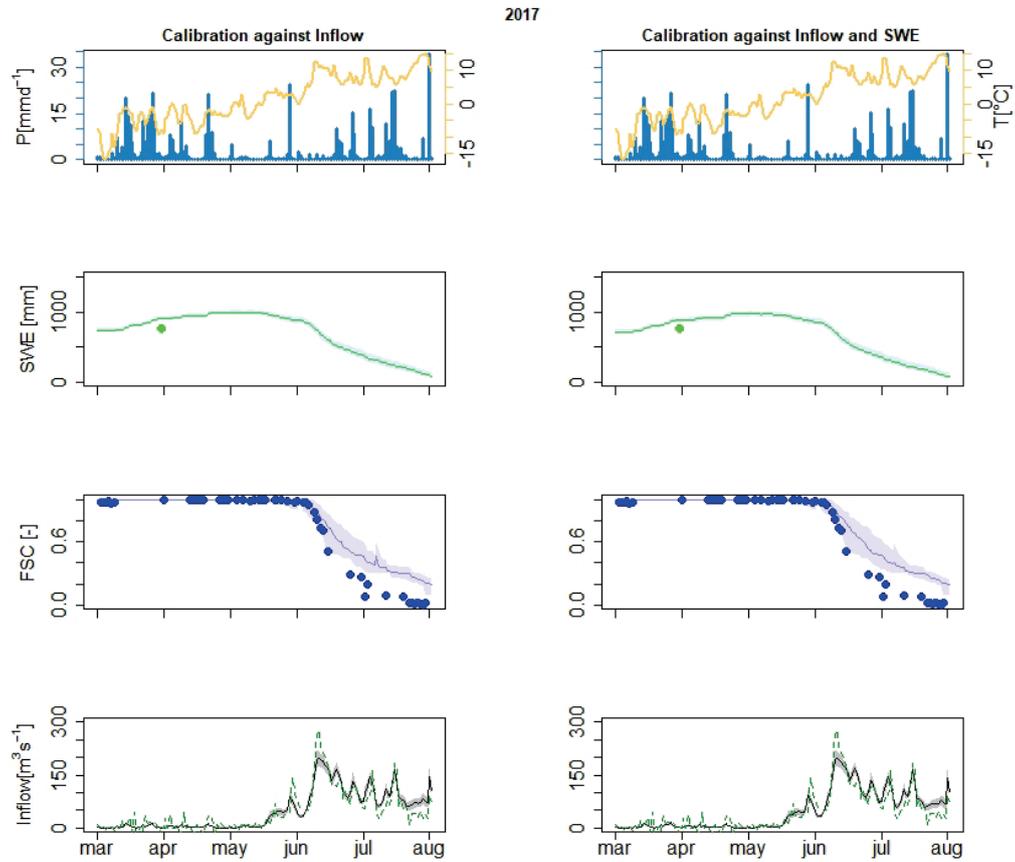


Figure 5. HYPE simulations by parameters calibrated against inflow,  $F1_Q$  and against inflow and SWE,  $F2_{Q,SWE}$  for the year 2017 during the calibration period. Green circles in the SWE subplots represent the SWE radar observations, blue circles in the FSC subplots represent the FSC observations, whereas the green dotted line in the local inflow subplots represent the observed local inflow. SWE, FSC and Inflow simulations within the 5th and 95th percentile uncertainty bounds together with their median are illustrated within their respective subplots.

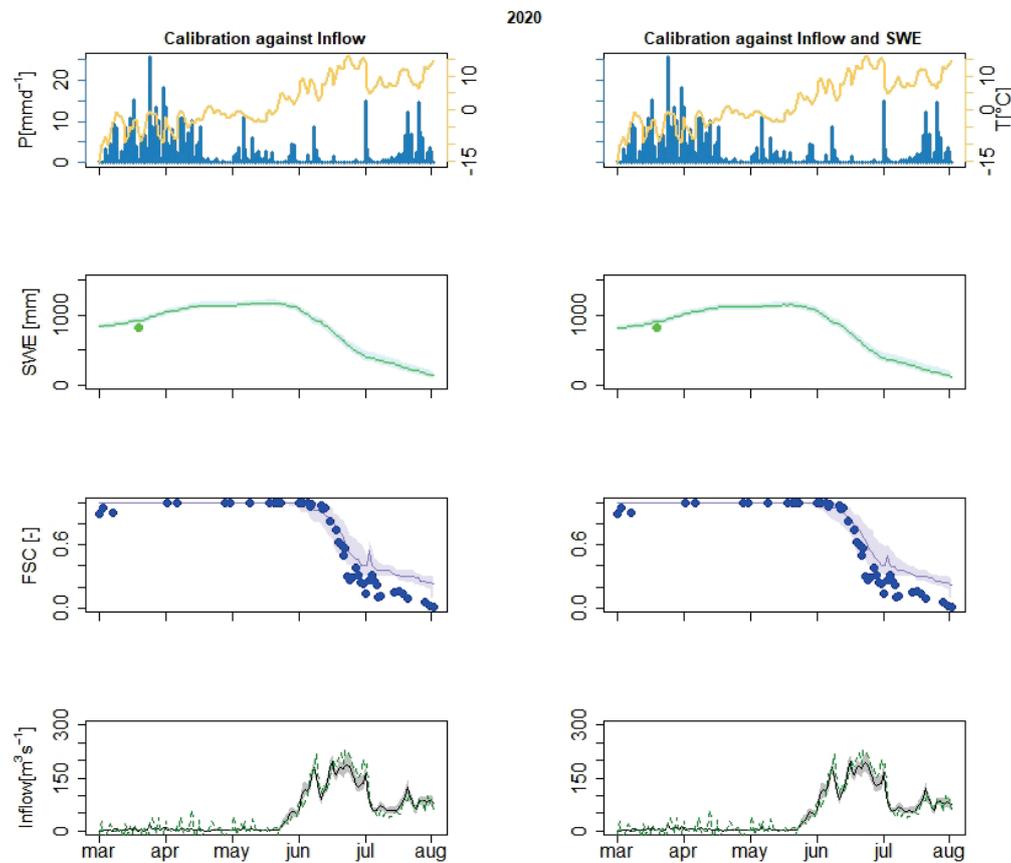


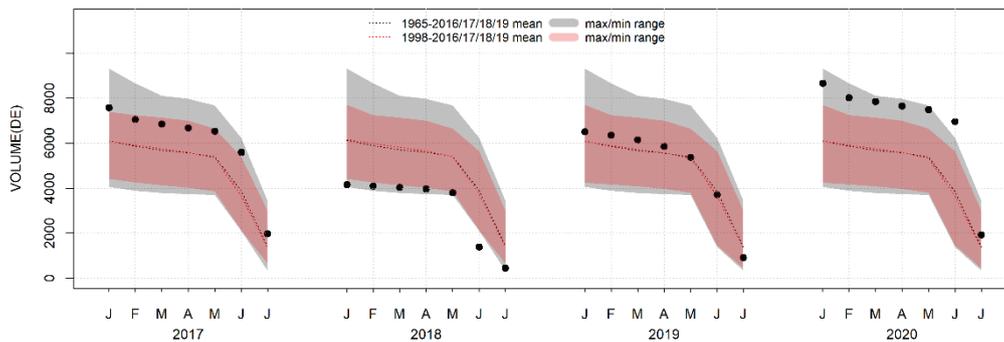
Figure 6. HYPE simulations by parameters calibrated against inflow,  $F1_Q$  and against inflow and SWE,  $F2_{Q,SWE}$  for the year 2020 during the verification period. Green circles in the SWE subplots represent the SWE radar observations, blue circles in the FSC subplots represent the FSC observations, whereas the green dotted line in the local inflow subplots represent the observed local inflow. SWE, FSC and Inflow simulations within the 5th and 95th percentile uncertainty bounds together with their median are illustrated within their respective subplots.

## 4.2 SPRING FLOOD FORECASTING AND DATA ASSIMILATION

The spring flood forecast experiments were conducted for the 4 years with snow survey data in Överuman 2017–2020. The study period contained a huge variation in snow conditions and spring inflow volumes. In fact, the observed inflows from 1st April–31st July in 2018 and 2020 were the 3rd lowest and 3rd highest, respectively, for the period for which inflow data was available (i.e. 1965–2020).

In the following analysis, the observed inflow from 1st of each month (January–July) until 31st July from all previous years since 1965 was used as a climatological reference forecast (Figure 12). When compared to the current observations, it can be noted that the observed inflows during 2018 and 2020 were close or outside of the previously recorded minimum and maximum, respectively, whereas the inflows during 2019 were very close to the climatological average throughout the winter and spring flood period. In 2017, the spring flood volume was approximately in between the average and the maximum of the previously recorded.

It could also be noted that the observed snow water equivalent from the VRF snow surveys were very well correlated with the spring flood volumes from 1st April–31st July. A simple linear model could predict the observed inflows with a mean absolute error lower than 5% (Figure 1 in Appendix C).



**Figure 7. Observed local inflow in Överuman from 1st of each month until 31st July for the months January–July 2017–2020 (black dots) compared to the climatological reference forecast based on all previous recorded inflow observations since 1965 (black dotted line/shaded area) and since 1998 (red dotted line/shaded area). Spring flood volumes are in day equivalent units (DE= 86,400 m<sup>3</sup>).**

#### 4.2.1 Impact on spring flood forecasts when using snow in calibration

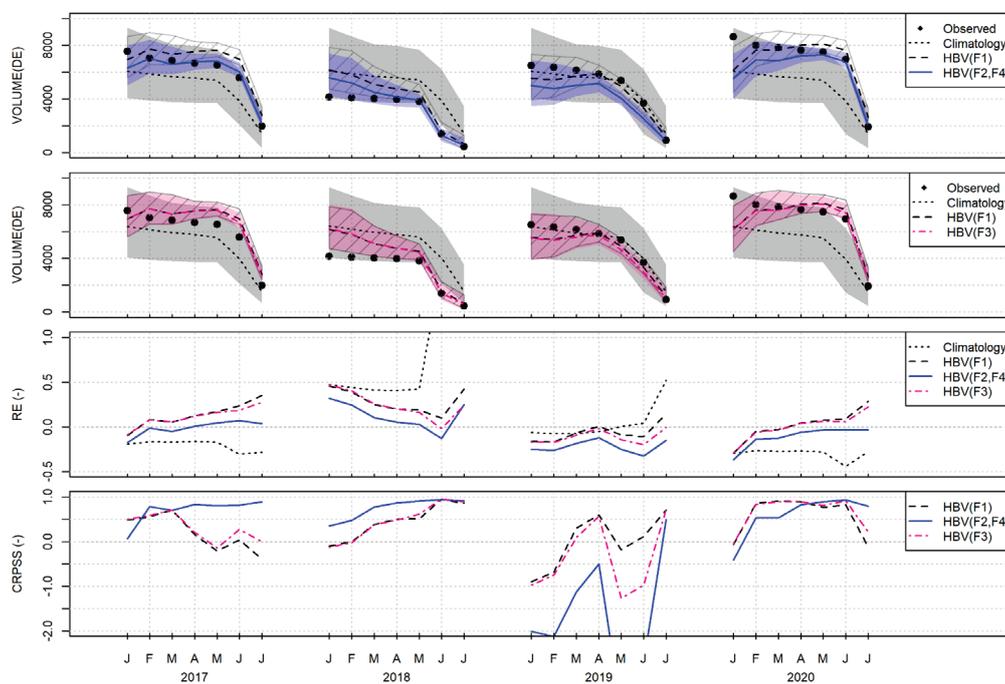
The impact of the different calibration strategies (with and without snow data) on the spring flood forecast performances are illustrated in Figures 13 and 14 for the HBV and HYPE models, respectively. Here we used the best parameter set from each calibration strategy to run the long-term forecasts in a deterministic way (i.e. without doing any correction of the model states before generating the forecasts).

In the HBV model, the largest positive impact on the spring flood forecasts was obtained when snow water equivalent was included in calibration (i.e. *F2*: inflow and SWE, and *F4*: inflow, SWE and FSC. Both were shown as one, HBV (*F2*, *F4*) in Figure 13 because they had the same optimal parameter). Based on the CRPSS scores, these models were consistently better than the other HBV models in 2017 and 2018, and equally good or better for the spring flood forecasts in 2020 between April and July.

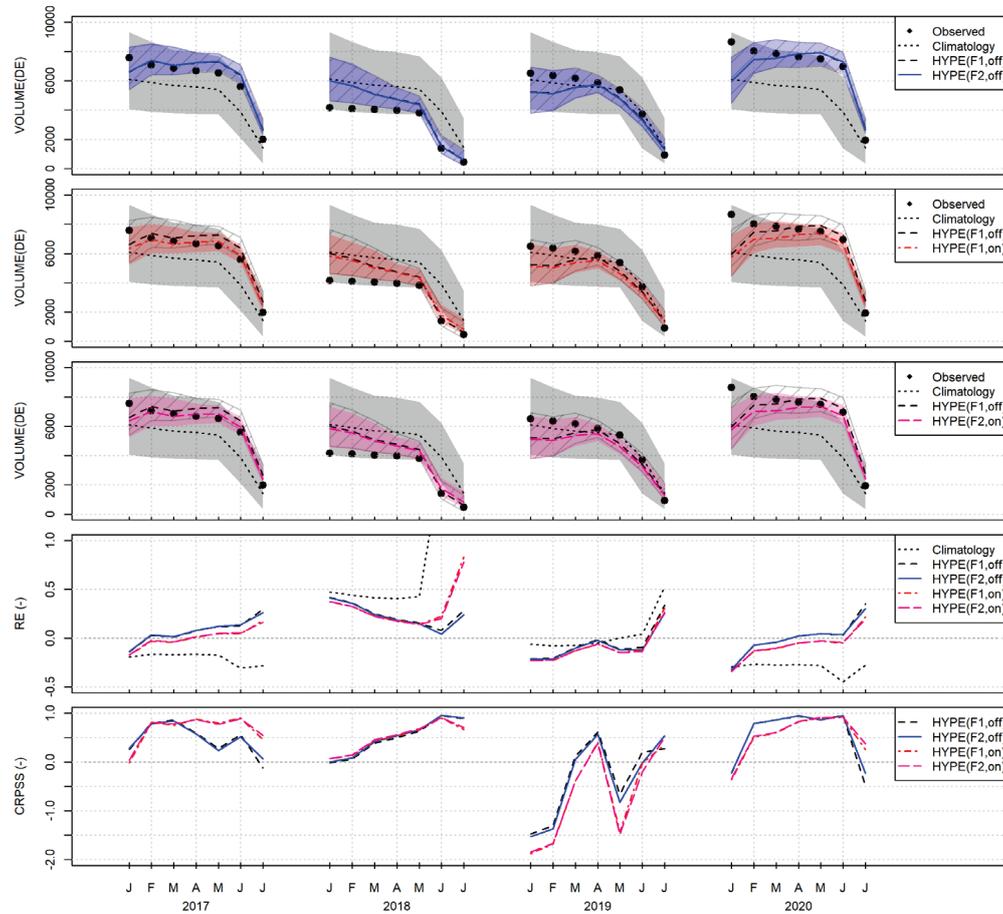
The calibration of HBV which included inflow and FSC (i.e. *F3*: inflow and FSC) provided small CRPSS improvements towards the end of the snow melt period in 2017 and 2020 compared to the one calibrated only against inflow (i.e. *F1*). Including other variables than inflow for the calibration of HBV resulted in a decreased forecast skill score during 2019.

Overall, it seems like the use of SWE in calibration of HBV reduced the snow storage in the HBV model, thus providing an improvement in the spring flood forecasts for the years with a general overestimation (2017, 2018 and 2020) and a decreased performance during the year with underestimated spring flood (2019). Similar analysis for the spring-flood forecasts resulting from using the current parameter set used in the HBV operational model can be found in Figure 2 of Appendix C.

In general, the impact of snow data calibration on the forecast skill was much smaller in the HYPE model compared to the impact in the HBV model. The main impact was seen by turning on and off the within-snow-class fractional snow cover parameterization. On the other hand, an improvement in the CRPSS scores can be seen towards the end of the snow melt period in 2017 and 2020, and a decrease forecast skill in 2019 for the HYPE model when the FSC function was on (similarly as it was seen when FSC was included in calibration for the HBV). The potential reasons for the lower sensitivity of the HYPE model to the snow data is further discussed in the discussion section.



**Figure 8.** Spring-flood inflow volume forecasts for Överuman 2017–2020 (deterministic models without data assimilation) using different calibration strategies to HBV. Simulations are compared to a climatological reference forecast (based on all observations since 1965) and to the observed inflow volumes for each respective year (black dots). Top two panels: HBV volume forecasts resulting from the calibration against only inflow, HBV (F1), against inflow, SWE and FSC, HBV (F2, F4), and against inflow and FSC, HBV (F3). Third panel: relative volume error (ensemble mean) for HBV forecasts shown in first two panels. Bottom panel: Continuous Ranked Probability Skill Score (CRPSS) for the HBV model forecasts, using the climatological forecasts as reference. Spring flood volumes for each month are defined as the total reservoir inflow volumes between the 1<sup>st</sup> of each month until the 31 July (in day equivalent units, DE= 86,400 m<sup>3</sup>). Mean, min and max values are shown for each ensemble.



**Figure 9.** Spring-flood inflow volume forecasts for Överuman 2017–2020 (deterministic models without data assimilation) using different calibration strategies to the HYPE model. Simulations are compared to a climatological reference forecast (based on all observations since 1965) and to the observed inflow volumes for each respective year (black dots). Top three panels: HYPE volume forecasts resulting from the calibration against only inflow, HYPE ( $F1, off$ ) and HYPE ( $F1, on$ ), and against inflow and SWE, HYPE ( $F2, off$ ) and HYPE ( $F2, on$ ). HYPE ( $Fn, off$ ) or HYPE ( $Fn, on$ ) refers to when the FSC function was on or off. Fourth panel: relative volume error (ensemble mean) for HYPE forecast ensembles shown in first three panels. Bottom panel: Continuous Ranked Probability Skill Score (CRPSS) for the HYPE model forecasts, using the climatological forecasts as reference. Spring flood volumes for each month are defined as the total reservoir inflow volumes between the 1<sup>st</sup> of each month until the 31 July (in day equivalent units, DE= 86,400 m<sup>3</sup>). Mean, min and max values are shown for each ensemble.

#### 4.2.2 Impact on spring flood forecasts when using inflow and snow in data assimilation

Finally, we assess the impact of using inflow and snow in data assimilation on the spring flood forecasts and forecast model initialization compared to the deterministic forecast initialization. Contrary to the previous section, inflow and snow data were used to correct the HBV and HYPE before producing the long-term spring flood forecasts.

In the case of the HBV model, model states were corrected using the auto-correction method, whereas snow storage was replaced with the SWE data at the date of observations. In the case of the HYPE model, different combinations of

inflow, SWE and FSC data were carried out using the EnKF method. Figures 6-10 in Appendix C illustrate the functionality of data assimilation for HYPE and HBV.

Table 7 shows a performance summary in terms of predictability of spring flood forecasts for the different calibration and data assimilation strategies.

**Table 7. Average Continuously Ranked Probability Skill Score (CRPSS) for spring flood forecasts issued 1st of April, May, and June for the different calibrations and data assimilation configurations of the HBV and HYPE models for the Överuman reservoir. Green color indicates when the forecast skill was improved by data assimilation compared to the deterministic initialization run with the same model/calibration method whereas the yellow color stands for the opposite. Underscore marks the forecast initialization with highest CRPSS for the particular model/calibration configuration, and bold underscore marks the forecasts with overall highest CRPSS scores. Curr stands for the current or established operational model at Överuman.**

Variables used in calibration	Label	Data-assimilation approach for forecast initialization								
		Det	Open	Q	SWE	Q SWE	FSC	Q FSC	Q SWE FSC	SWE FSC
<b>HBV</b>										
Q (Curr)	<i>Curr</i>	-0,50		<i>-0,67</i>	<u>0,52</u>	0,50				
Q	<i>F1</i>	0,39		0,65	0,65	<u>0,72</u>				
Q, SWE	<i>F2,F4</i>	0,14		0,43	0,53	<u>0,67</u>				
Q, SWE, FSC										
Q, FSC	<i>F3</i>	0,28		0,59	0,61	<u>0,72</u>				
<b>HYPE</b>										
Q	<i>F1,off</i>	0,54	<i>0,52</i>	0,58	0,66	<u>0,72</u>	0,57	0,55	0,70	0,68
Q, SWE	<i>F2,off</i>	0,50	<i>0,49</i>	0,61	0,68	<u>0,73</u>	0,52	0,60	<u>0,73</u>	0,68
Q, SWE, FSC	<i>F2,on</i>	0,55	<i>0,45</i>	<i>0,54</i>	<i>0,45</i>	0,55	<u>0,64</u>	0,60	0,56	0,57
Q, FSC	<i>F1,on</i>	0,54	<i>0,46</i>	0,58	<i>0,51</i>	0,60	<u>0,63</u>	<u>0,63</u>	0,60	0,61

The summary shown in Table 7 is based on averaging CRPSS over the forecasts issued in April–June. Certainly, the selection of months for computing the average CRPSS scores plays a role as this is a quite sensitive measure but we chose the 3 months when data assimilation of inflow and SWE should have the largest impacts on the spring-flood forecast volumes, and before the main bulk of the snow melt runoff that generally happened during June.

Results showed that assimilation of inflow (Q), snow water equivalent (SWE), and/or fractional snow cover data (FSC), individually or in combination, generally improved the spring flood forecasts for almost all combinations of model/calibration/initialization methods (even the forecasts using the current parameter set in HBV operational model). The highest CRPSS scores were achieved

by assimilation of Q and SWE data in combination, with similar scores (0,72-0,73) for the HBV and HYPE models.

In the case of HBV, the highest CRPSS scores (0,72) were achieved assimilating both Q and SWE (i.e. *F1* and *F3* models). The HBV models calibrated against SWE (i.e. *F2*, *F4*) achieved slightly lower average CRPSS with all tested initialization methods compared to the other models (i.e. *F1*, *F3*), but they still achieved a high CRPSS score (i.e. 0,67) when both Q and SWE were used for assimilation. Furthermore, the latter models calibrated against SWE achieved the largest improvements in CRPSS from the deterministic initialization (i.e. CRPSS improved from 0,14 to 0,67).

In the case of HYPE, the highest CRPSS scores (0,70-0,73) were achieved by assimilating either Q and SWE in combination, or by assimilating Q, SWE and FSC in combination (i.e. *F1, off* and *F2, off* models). The HYPE model calibrated with SWE (*F2, off*) had a slightly higher increase in CRPSS between deterministic and assimilation initialization, compared to the model calibrated only with inflow. The HYPE models with the fractional snow cover function enabled (i.e. *F1, on* and *F2, on*) improved the most when assimilating FSC data.

The absolute improvements in the CRPSS April–June by data assimilation compared to the deterministic forecast initialization was larger for the HBV models than for the HYPE models, but mainly because the deterministic HYPE models had better CRPSS scores than to the HBV models. The CRPSS scores for individual forecast months and years for all combinations of model/calibration/initialization methods are shown in Figures 3–5 in Appendix C.

## 5 Discussion

The aim of this report was to assess the ability of the hydrological models HBV and HYPE to describe snow accumulation, snow melt and spring flows, and to explore the integration of different types of snow measurements in model calibration and forecast model initialization.

The data assimilation experiments showed clearly that including snow information in forecast model initialization improved the forecast skill with regard to spring flood runoff volume, which provides a strong argument for the added value of snow observation for hydropower reservoir management. Almost independent of model structure or calibration strategy, assimilation of the snow water equivalent data improved the spring flood forecast performance. This was clearly seen when SWE was used together with inflow for data assimilation in both models (i.e. this was when forecast performance improved the most). Including fractional snow cover in data assimilation in HYPE provided mixed results as the forecast performances improved or decreased depending on how the representation of snow distribution was set up in the model. Between the two types of snow data used for data assimilation, SWE provided the largest value for improving forecast performance.

These results are also in agreement with previous findings by Gustafsson et al (2015) and Muusuza et al (2020), which showed improvements in spring flood forecasting by assimilation of inflow data in combination with fractional snow cover from satellite, and similar type of in-situ based snow water equivalent data. The main difference from the previous studies is that the improvements in forecast skills by snow data assimilation was more consistent in the present results. Unpublished results by Björn Norell, VRF Östersund (personal communication) also confirms the present results with large improvements in the spring flood forecasts by assimilation of the Överuman SWE observations using the established and calibrated HBV model.

Regarding the choice of data assimilation method, it could be noted that the HBV auto-correction method and the HYPE EnKF method in some senses operate in a similar way. They both adjust the meteorological forcing variables (temperature and precipitation) to improve the agreement between simulated and observed runoff while adjusting the hydrological state variables. In theory, the auto-correction method is more similar to a particle filter, where model is driven by different perturbations on precipitation and temperature that are either accepted or rejected until an optimal selection of perturbations are found, whereas the EnKF method scales the model state vector based on co-variances between model state errors and the errors in the observed variables.

A key element of the EnKF method is that it is able to assimilate any type of observation which the model is able to predict, and that several types of observations can be assimilated simultaneously in a consistent framework. A weakness of the EnKF method is related to the basic assumption of normally distributed errors and linear relationships between model error co-variances and model-state errors which are not usually the case. These inconsistencies may

reduce the efficiency of the EnKF data assimilation if not adjusted properly. It is still a matter of research and development to fine-tune the implementation and the applications of the EnKF in the HYPE model, but the results in this report show that it is getting closer to a stable product. It should also be noted that the EnKF is implemented as an independent Fortran90 module within the HYPE model framework and could be adapted in any other similar hydrological model framework.

On the other hand, including snow information in the model calibration did not show any improvement in the model robustness from calibration to verification periods with regard to runoff. Adding snow observations in the calibration would likely reduce the performance with regard to runoff but at the same time improves the robustness from calibration to verification period, and thus avoid getting the right answer for the wrong reason. Why this was not demonstrated in our calibration experiments may be due to the shorter time period with snow observations, or too much weight to the runoff in the calibration procedure. But it may also be an indication of limitations in the model structure to represent the relation between meteorological conditions, snow and runoff in the different climate conditions of the calibration and verification periods.

When comparing the information content between the inflow and snow data, it is unlikely that using only two years of SWE data in calibration can represent all the variation in weather conditions as represented by the inflow data. However, including small amount of SWE data in calibration can still provide value for rejecting models or parameter sets that largely overestimate or underestimate SWE measurements making the models more realistic internally. Here, parameter sets were rejected if they overestimated or underestimated SWE by more than 30%.

The total spring melt runoff volume is maybe not only a question of the snow melt, but rather an integrated effect of all snowfall events during the winter, and temperature increase and rainfall during the spring. The low impact in both models when including SWE data in calibration could also be related to the temporal sampling of this data (i.e. one measurement sampled about yearly intervals). Previous research has reported that continuous measurements of individual events may provide sufficient and more information for model calibration than measurements sampled on fixed intervals (McIntyre & Wheeler, 2004, Seibert & McDonnell, 2013). Calibration-data needs rely on the information content of the observations rather than on the pure number of individual observations. This was shown in a scenario of data scarcity by Reynolds et al. (2020). There they compared different scenarios of data availability of continuous measurements in calibration and their results showed that relatively equal information content in terms of predictability was obtained after the model was calibrated with three or more events. The latter may apply as well for snow-data needs in calibration. It is assumed that the value of including snow data in calibration will increase if more than one measurement per winter could be included for such procedure, as well as by extending the number of years with similar SWE observations as those used in this study (i.e. close to the time of annual maximum accumulation).

As it happened in data assimilation, SWE was more valuable or informative in calibration than FSC for improving both models' ability for describing snow accumulation and snow melt. This could have been caused by the weighting factor value used for them when computing model performance (i.e. 0,3 for both). Such weighing factor value had a greater impact when integrating SWE in calibration than it did for FSC. A greater value for the weighting factor value of FSC (i.e.  $W_{FSC}$ ) would have caused a bigger impact in the selection of the parameters and would have likely improved FSC predictability. Of course, a greater weighting factor for SWE than the one used here would have a similar effect. Another more robust objective function for measuring the predictability of FSC rather than the Pearson correlation could have caused a larger impact in calibration too, as high correlations in model performance did not result in good fitting of the FSC observations (i.e. both models failed to reproduce FSC towards the end of snow melt period).

In terms of model performance, all the calibrations tested in this experiment outperformed the HBV parameter set currently used in the operational model. The latter comes without a surprise since that model was calibrated for a period before changes to the recording station at Överuman were made. After the Klippen's power station was put in operation, the forecasts at Överuman with the current model began to overestimate (about 30%). Back then, these errors were assumed to be related to errors on fall losses not considered and could not be explained by the climatological input data. Curiously, the established and calibrated model at Överuman increments precipitation 30% (i.e.  $p_{corr} = 1,3$ ) which now result in large overestimations of inflow and SWE volume. As of now, VRF and SMHI are well aware that the operational model returns simulations that does not agree with the discharge observations from the Klippen's power station and the reliability of the latter has still to be determined. The calibration procedures carried out in this experiment used a calibration period based on recordings by the Klippen's power station which resulted in a good representation of snow accumulation, snow melt and spring flows. It would be desirable to determine the reliability of the observations from the Klippen's power station in order to determine if an updating of the parameter set used in the operational model is required. In the assumption that such data is reliable, the current operational model requires large corrections before producing short- and long-term forecasts.

Many hydrological processes in HYPE and HBV are represented in a similar way. The main difference between the two models is in the way they represent snowfall distribution: for HYPE this is based on current wind direction and physical characteristics, whereas for HBV model this is based on the empirical distribution of the parameters. Overall, both models performed relatively similar with regards their ability to describe snow accumulation, snow melt and spring flows with their semi-distributed spatial representation of the processes. When both models were calibrated only against inflow, HYPE showed a slightly better predictability of SWE than HBV but after including SWE in calibration these differences in performance leveled off. To some extent, HBV is more dependent on including snow data in calibration to achieve good fits in snow storage than HYPE does, most likely because of how they represent snow distribution. It is assumed that the value added on HYPE by representing snowfall distribution based on wind

direction will be more evident when transferring the calibrated model to areas without snow data, since the calibrated parameters are related to the topographic analysis and wind direction data.

Several parameter sets were accepted as representative in calibration rather than only using only one as optimal parameter. Here we reject the concept of an optimal parameter set in favor of accepting multiple parameter sets that give acceptable results and that are nearly as good as the optimal one (i.e. equifinality). It is very unlikely that an optimal parameter set for a calibration period is the optimum for another period and accepting multiple parameter sets as good is a way to deal with this issue. Although in our study this concept was only implemented in the calibration experiments tested here there is no impediment on using multiple parameter sets as well to generate long-term spring flood forecasts and look into the ensemble uncertainty generated by the parameter uncertainty.

## 6 Conclusions

The overall aim of this study was to explore how different types of snow measurements can be integrated and used in hydrological modelling to better describe spring flows, snow accumulation and snow melt. This was done in two steps. The first part of the experiment was designed to explore the value of including snow measurements during the calibration procedure whereas the second part of the experiment was to assess how long-term spring flow forecasts improves if we previously include snow information during the data assimilation procedure. Our methodology was tested in the HBV- and HYPE-model versions currently in used by SMHI. The specific conclusions from our analysis are given below:

1. When snow data was included or excluded in calibration model performance was relatively similar for both models with regards their ability to describe snow accumulation, snow melt and spring flows. However, the results did indicate any impact on the robustness of the model performance between calibration and verification periods when snow data was included in calibration.
2. The added value of the snow data in terms of improved forecast performance was mainly seen when the snow survey data was assimilated during forecast model initializations.
3. Assimilation of reservoir inflow, snow water equivalent, and fractional snow cover data individually or in combination during HBV and HYPE forecast model initialization generally improved the long-term spring flow forecast for all years and forecast periods from April-July.
4. The combinations of assimilating both inflow and snow water equivalent data provided the highest forecast skill scores.
5. Models with lower forecast skill using a deterministic initialization (without data assimilation), in most cases achieved a larger improvement in forecast skill when assimilation of inflow and snow data was used in the initialization.
6. The data assimilation method used in the HBV model resulted in a larger improvement in the forecast skill score than the data assimilation method used in HYPE, but the highest forecast skill scores were similar in the two models since the calibrated HYPE models had a slightly higher forecast skill score without data assimilation.

## 7 Follow-up

Finally, it must be emphasized that the results found herein are only valid for one basin and two model structures. The generality of our results can only be proven when the methodology is tested on other basins and models.

As a follow-up to our study, the following experiments could be made to further assess the ability of including snow data in calibration and data assimilation in HBV and HYPE:

1. Repeating the calibration experiments carried out in this study but with a calibration period of only two (or four) years to have same number of years with both snow and inflow data. Additionally, the weighting factors for including snow data in calibration should have larger values than those used here to give more weight on fitting snow data while still providing good fit to the inflow data.
2. Testing our methodology in other basins where snow data is available. Kultsjön within the Ångermanälven river basin could be a possible study site given that it counts with such data. Snow data has been previously included in the operational model for Kultsjön and has showed improvements in forecast performance.
3. Adapting the models so that the continuous measurements of snow-depth data from surrounding stations can be included. This would allow to assess impact of temporal sampling of SWE. Daily snow-depth measurements at the SMHI stations in Mjölkbäcken and Hemavan are available which were not used in this study (the former station is located inside the Överuman basin). Moreover, there are a handful of stations in Norway within 10 km of the study area which can also be used. The challenge for using data from these stations are to transform either the modelled SWE to represent the local snow depth at the station location, or the other way around. This would require that snow densification is represented in the models, which is the case for the HYPE model or a distributed model approach.
4. Further development of the data-assimilation method implemented in HBV. Including other methods such as EnKF would allow updating or corrections of different types of observations simultaneously (i.e. inflow and SWE).
5. Including parameter uncertainty not only calibration but also for generating the long-term spring flood forecasts after snow data has been included in the models.

## 8 References

- Bergström, S. (1976) Development and application of a conceptual runoff model for Scandinavian catchments. SMHI, Report No. RHO 7. Norrköping, Sweden.
- Freudiger, D., Kohn, I., Seibert, J., Stahl, K. & Weiler, M. (2017) Snow redistribution for the hydrological modeling of alpine catchments. *WIREs Water* 4(5). doi:<https://doi.org/10.1002/wat2.1232>
- Gustafsson, D., Lindström, G. & Kuentz, A. (2015) Vårflodsprognoser med snöuppdatering. Energiforsk rapport 2015:208.
- Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15(5), 559–570.
- Hägemark, Lars & Ivarsson, Karl-Ivar & Gollvik, Stefan & Olofsson, Per-Olof. (2003). MESAN, an operational mesoscale analysis system. *Tellus A*. 52. 2 - 20. [10.1034/j.1600-0870.2000.52102.x](https://doi.org/10.1034/j.1600-0870.2000.52102.x).
- Johansson, B. (2000) Areal precipitation and temperature in the Swedish mountains. An evaluation from a hydrological perspective. *Nord. Hydrol.* 31, 207–228.
- Johansson, B. & Chen, D. (2003) The influence of wind and topography on precipitation distribution in Sweden: Statistical analysis and modelling. *Int. J. Climatol.* 23, 1523–1535.
- Johansson, Barbro, Dahmé, J. & Gustafsson, D. (2015) Vindrelaterad snöfördelning i hydrologiska modeller. Energiforsk rapport 2015:207.
- Lindström, G., Johansson, B., Persson, M., Gardelin, M. & Bergström, S. (1997) Development and test of the distributed HBV-96 hydrological model. *J. Hydrol.* 201(1–4), 272–288. doi:[10.1016/S0022-1694\(97\)00041-3](https://doi.org/10.1016/S0022-1694(97)00041-3)
- Lindström, G., Pers, C., Rosberg, J., Strömquist, J., and Arheimer, B.: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, *Hydrology Research*, 41, 295–319, <https://doi.org/10.2166/nh.2010.007,2010>.
- McIntyre, N. R. & Wheeler, H. S. (2004) Calibration of an in-river phosphorus model: Prior evaluation of data needs and model uncertainty. *J. Hydrol.* 290(1–2), 100–116. doi:[10.1016/j.jhydrol.2003.12.003](https://doi.org/10.1016/j.jhydrol.2003.12.003)
- Musuza, J.L.; Gustafsson, D.; Pimentel, R.; Crochemore, L.; Pechlivanidis, I. Impact of Satellite and In Situ Data Assimilation on Hydrological Predictions. *Remote Sens.* 2020, 12, 811.
- Reynolds, J. E., Halldin, S., Seibert, J., Xu, C. Y. & Grabs, T. (2020) Flood prediction using parameters calibrated on limited discharge data and uncertain rainfall scenarios. *Hydrol. Sci. J.* 65(9), 1512–1524. doi:[10.1080/02626667.2020.1747619](https://doi.org/10.1080/02626667.2020.1747619)
- Reynolds, J. E., Halldin, S., Xu, C. Y., Seibert, J. & Kauffeldt, A. (2017) Sub-daily runoff predictions using parameters calibrated on the basis of data with a daily temporal resolution. *J. Hydrol.* 550, 399–411. doi:[10.1016/j.jhydrol.2017.05.012](https://doi.org/10.1016/j.jhydrol.2017.05.012)

- Samuelsson, P., Gollvik, S., and Ullerstig, A.: The land-surface scheme of the Rossby Centre regional atmospheric climate model (RCA3), Tech. rep., SMHI, Norrköping and Sweden, 2006.
- Schwaizer, G., E. Ripper, T. Nagler, J. Ikonen, K. Luojus, T. Ryyppö, M. Hiltunen, J. Pulliainen and S. Metsämäki. 2017. CryoLand, Copernicus Service Snow and Land Ice - Upgraded Pan-European Snow Services.  
[http://cryoland.enveo.at/downloads/CryoLand\\_No262925\\_ID4.6\\_IUSSv10\\_v1.1A.pdf](http://cryoland.enveo.at/downloads/CryoLand_No262925_ID4.6_IUSSv10_v1.1A.pdf)
- Seibert, J & McDonnell, J. J. (2013) Gauging the Ungauged Basin: The Relative Value of Soft and Hard Data. *J. Hydrol. Eng.* **20**(1).  
 doi:10.1061/(ASCE)1084-0699(2013)18:1(1)
- Seibert, J. (1997) Estimation of Parameter Uncertainty in the HBV Model. *Nord. Hydrol.* **28** (4/5), 247–262.
- Seibert, J. (1999) Regionalisation of parameters for a conceptual rainfall-runoff model. *Agric. For. Meteorol.* **98–99**, 279–293. doi:10.1016/S0168-1923(99)00105-7
- SMHI & UU. (2020) SNODDAS (SNOW Distribution and Data Assimilation for improved spring flood forecast and Sustainable hydropower reservoir regulation. Retrieved from  
<https://www.geo.uu.se/research/luval/disciplines/Physical-Geography/ongoing-research/snoddas/>
- Winstal, A., Elder, K., and Davis, R. E.: Spatial Snow Modeling of Wind-Redistributed Snow Using Terrain-Based Parameters, *Journal of Hydrometeorology*, 3, 524–538, [https://doi.org/10.1175/1525-7541\(2002\)003<0524:SSMOWR>2.0.CO;2](https://doi.org/10.1175/1525-7541(2002)003<0524:SSMOWR>2.0.CO;2), 2002.

## Appendix A: Stage-volume table for Lake Överuman and SWE data at study site.

Table 1. Stage level and volume table for Lake Överuman.

Stage levels (m – RH200)	Volume (m <sup>3</sup> )
0,0	0,0
520,2	0,1
520,5	253,0
521,0	683,0
521,5	1128,0
522,0	1589,0
522,5	2068,0
523,0	2548,0
523,5	3040,0
524,0	3532,0
524,5	4034,0
524,7	4234,0
525,0	4540,0
525,5	5044,0

Table 2. Estimates of snow water equivalent (SWE) at Överuman.

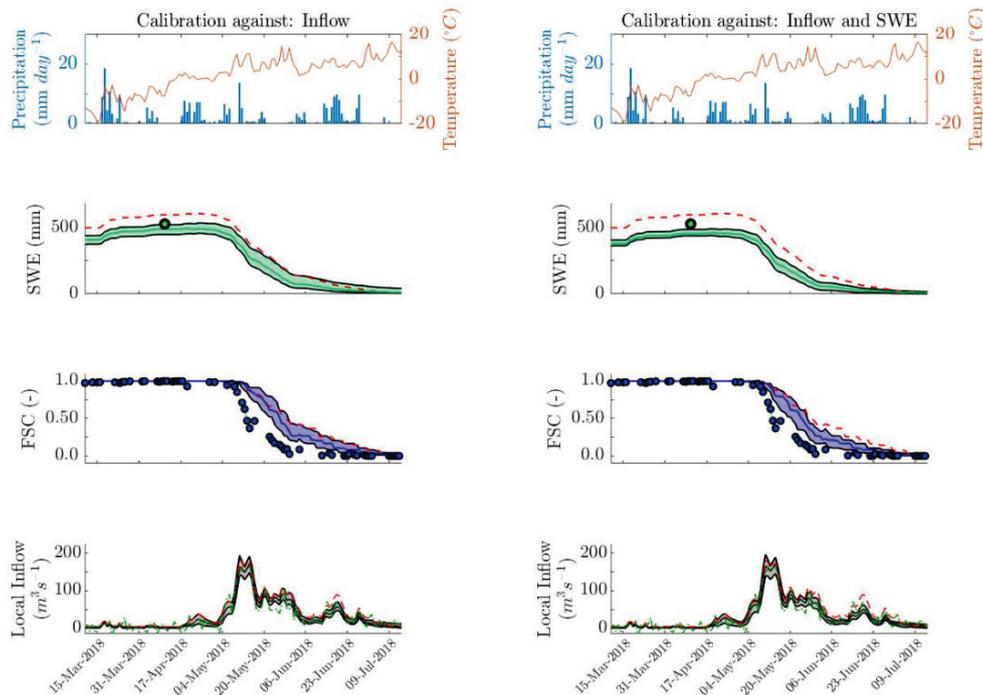
Date	SWE (mm)
2017-03-30	763
2018-04-11	526
2019-03-28	631
2020-03-19	813

## Appendix B: Additional results on model calibration when integrating snow data

Table 1 shows the numbers of parameter sets in common for the different calibrations tested using HBV. Figures 1–4 compare HBV and HYPE model results by different calibrations tested during calibration and verification.

**Table 1. Numbers of parameter sets in common for the different calibration approaches of the HBV model.**

	$F1_Q$	$F2_{Q,SWE}$	$F3_{Q,FSC}$	$F4_{Q,SWE,FSC}$
$F1_Q$	-	46	75	37
$F2_{Q,SWE}$	46	-	50	86
$F3_{Q,FSC}$	75	50	-	46
$F4_{Q,SWE,FSC}$	37	86	46	-



**Figure 1. HBV simulations by parameter sets calibrated against inflow,  $F1_Q$  and against inflow and SWE,  $F2_{Q,SWE}$  for the year 2018 during the calibration period. Green circles in the SWE subplots represent the SWE radar observations, blue circles in the FSC subplots represent the FSC observations, whereas the green dotted line in the local inflow subplots represent the observed local inflow. For all subplots the read dotted lines represent the simulations obtained for the current parameter set used in the operational model. Local inflow, SWE and FSC simulations within the 5th and 95th percentile uncertainty bounds together with their median are illustrated within their respective subplots.**

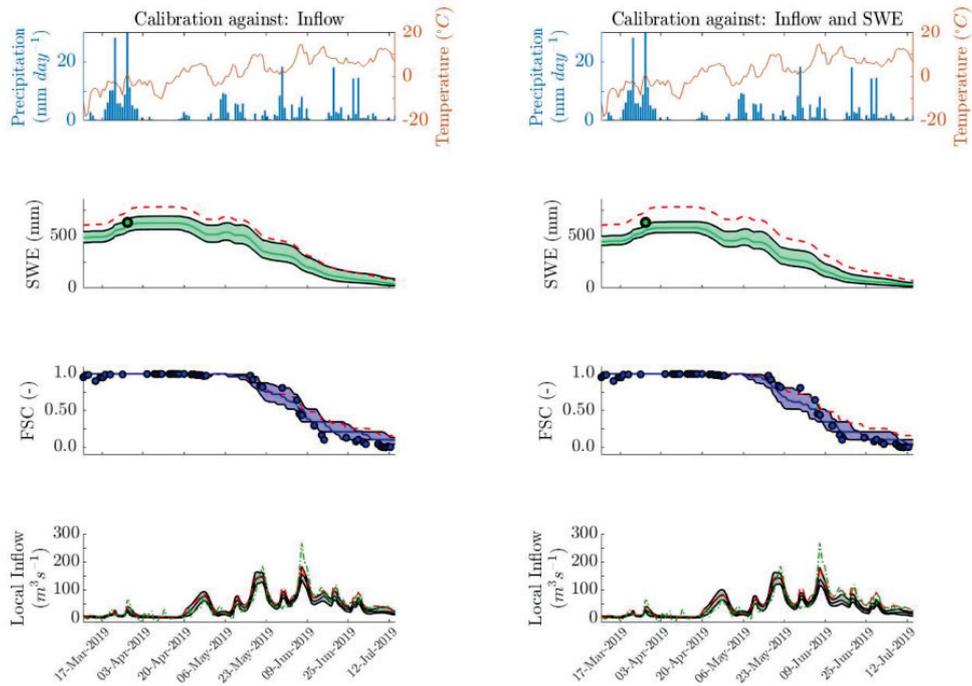
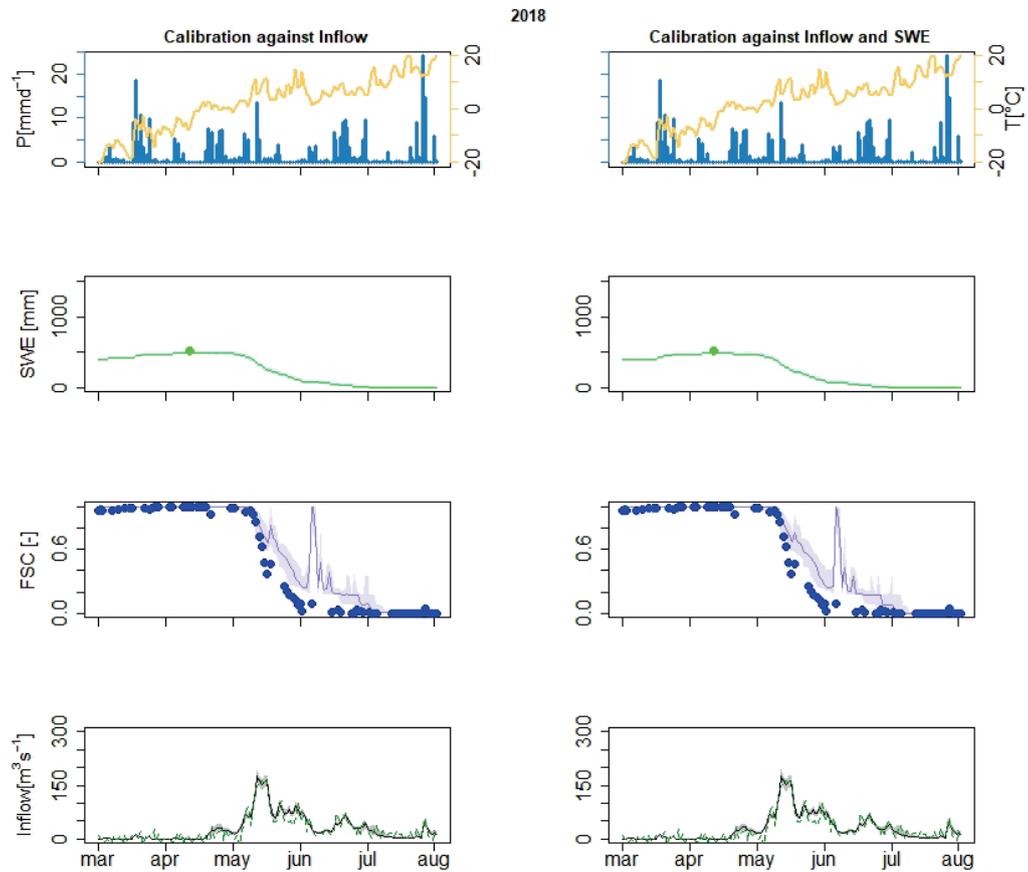


Figure 2. HBV simulation by parameter sets calibrated against inflow,  $F1_Q$  and against inflow and SWE,  $F2_{Q,SWE}$  for the year 2019 during the verification period. Green circles in the SWE subplots represent the SWE radar observations, blue circles in the FSC subplots represent the FSC observations, whereas the green dotted line in the local inflow subplots represent the observed local inflow. For all subplots the red dotted lines represent the simulations obtained for the current parameter set used in the operational model. Local inflow, SWE and FSC simulations within the 5th and 95th percentile uncertainty bounds together with their median are illustrated within their respective subplots.



**Figure 3.** HYPE simulations by parameter sets calibrated against inflow,  $F1_Q$  and against inflow and SWE,  $F2_{Q,SWE}$  for the year 2018 of the calibration period. Green circles in the SWE subplots represent the SWE radar observations, blue circles in the FSC subplots represent the FSC observations, whereas the green dotted line in the local inflow subplots represent the observed local inflow. SWE, FSC and Inflow simulations within the 5th and 95th percentile uncertainty bounds together with their median are illustrated within their respective subplots.

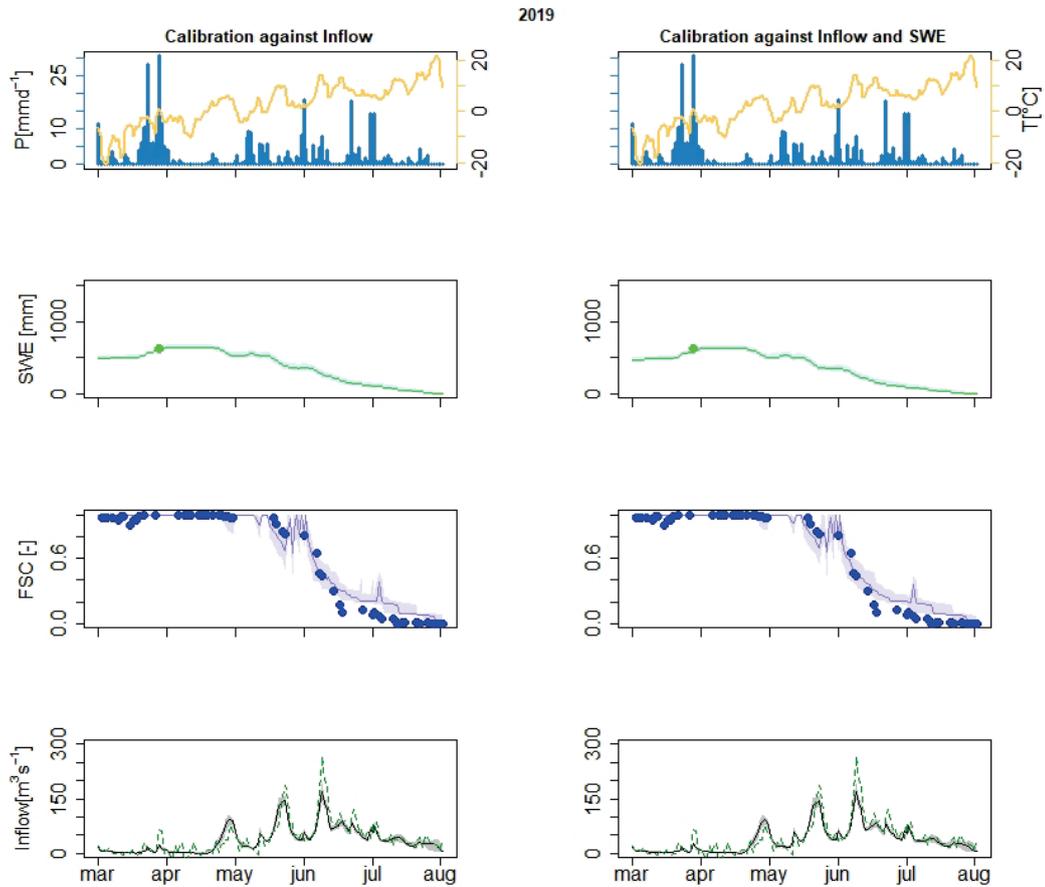


Figure 4. HYPE simulations by parameter sets calibrated against inflow,  $F1_Q$  and against inflow and SWE,  $F2_{Q,SWE}$  for the year 2019 of the calibration period. Green circles in the SWE subplots represent the SWE radar observations, blue circles in the FSC subplots represent the FSC observations, whereas the green dotted line in the local inflow subplots represent the observed local inflow. SWE, FSC and Inflow simulations within the 5th and 95th percentile uncertainty bounds together with their median are illustrated within their respective subplots.

## Appendix C: Additional results on data assimilation and long-term spring flood forecasts

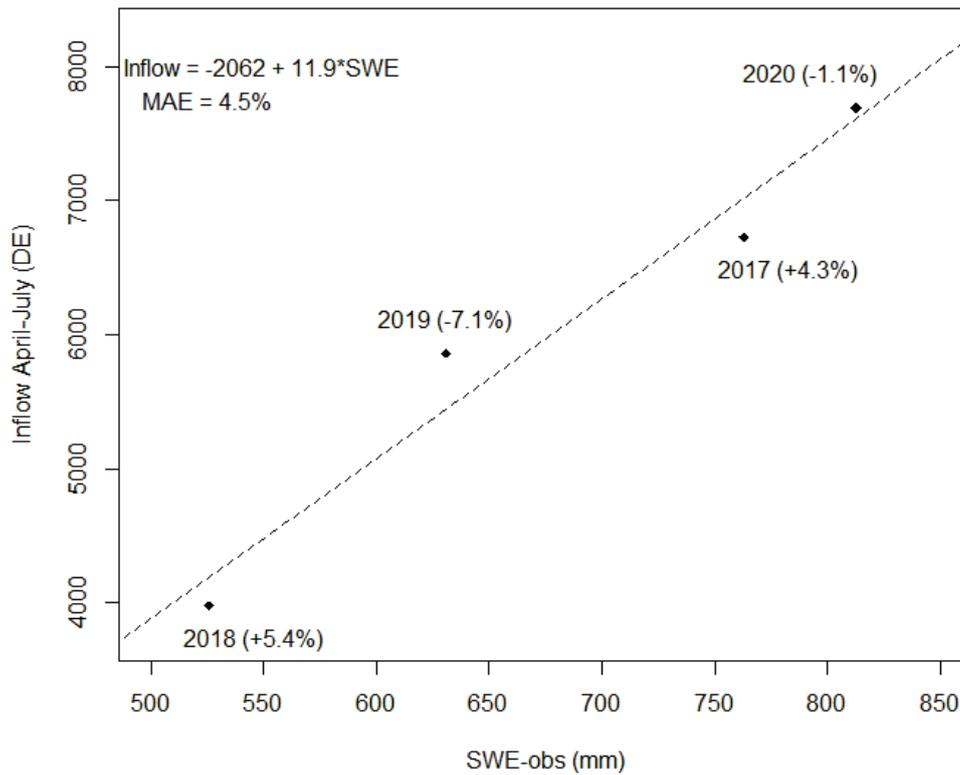
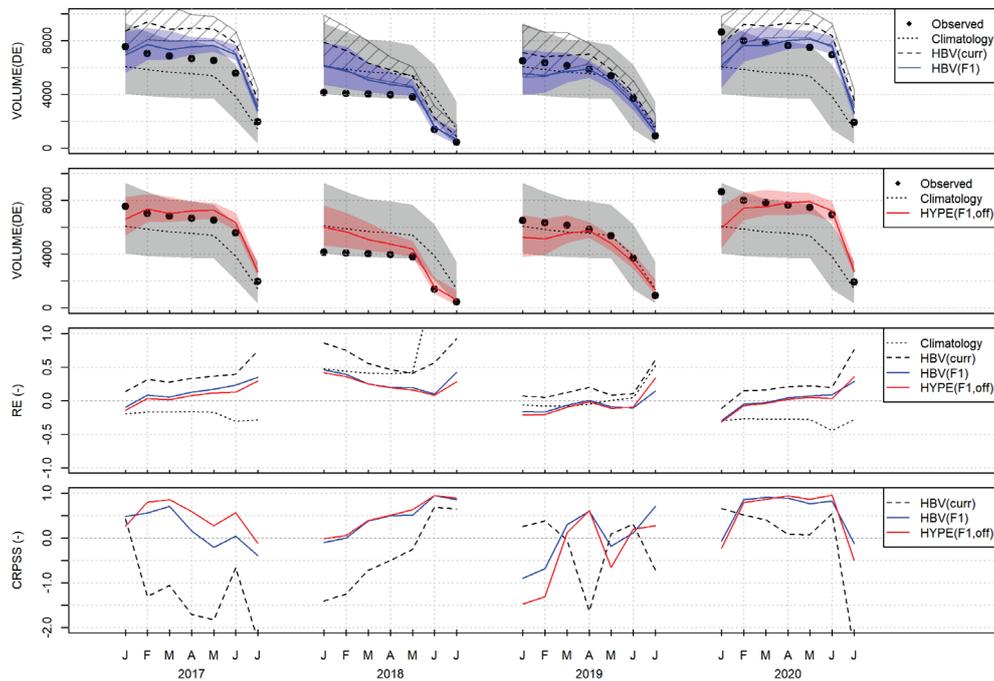


Figure 1. Observed reservoir inflow 1st April-31st July plotted versus the snow water equivalent estimated from the snow surveys in the Överuman basin 2017–2020 (dots), where the line is a simple linear regression model fitted to the 4 data points. Spring flood volumes are in day equivalent units (DE= 86,400 m<sup>3</sup>). The relative volume error for each year is show in brackets, and the mean absolute error (MAE) is shown in the legend.

Figure 2 shows the predictive ability of the spring-flood forecasts of the current parameter set used in the HBV operational model with respect to the climatological reference model (1965–2021), together with the HBV and HYPE models calibrated only against inflow in this experiment. Here, the spring-flood forecasts were generated deterministically (i.e. without data assimilation) and the results are summarized as follow:

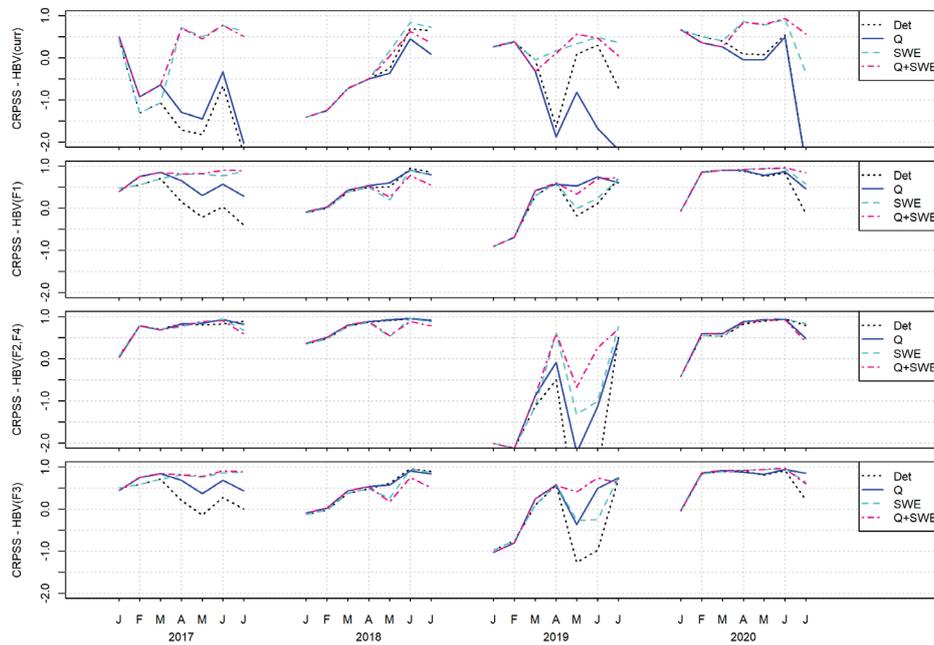
- Forecasts using the calibrated HBV( $F1$ ) and HYPE( $F1,off$ ) models were clearly better than both the climatological reference forecasts and the current HBV model.
- The relative errors of the calibrated HBV and HYPE models were distributed around 0 and their mean errors was clearly smaller than the climatological reference model, except for the unusually normal year 2019.
- The HBV( $F1$ ) and HYPE ( $F1, off$ ) models also had a continuous ranked probability skill score (CRPSS) larger than 0 which shows that they provided a clear improvement compared to the reference model. The CRPSS of the calibrated models were occasionally below 0 even though the ensemble mean relative error was smaller or similar to the error in the reference forecast. An interpretation of this might be that the HBV and HYPE forecast ensembles were over-confident (too small variance) and the observations actually fell too far outside the minimum-maximum range of the forecasts. An example of this can be seen for the HBV and HYPE forecasts in May–July 2017, and January and July 2020.
- Unsurprisingly, the current HBV model forecasts systematically overestimated the observed inflow volumes and their CRPSS values were typically below 0, meaning that these forecasts were of less quality than the climatological reference model.



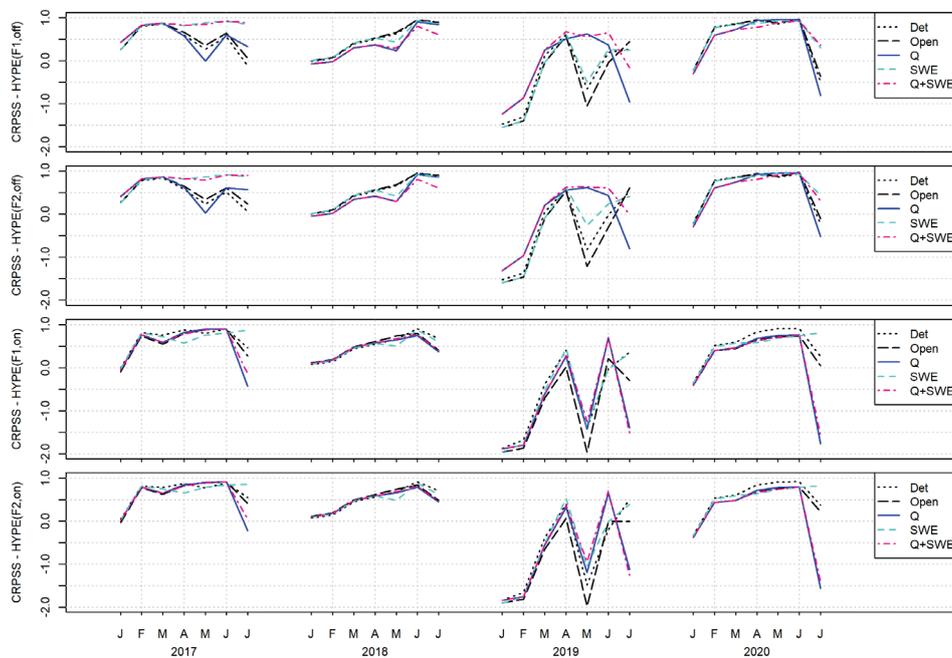
**Figure 2. Spring-flood inflow volume forecasts for Överuman 2017–2020 (deterministic models without data assimilation) using the HBV (*F1*) and HYPE (*F1, off*) models compared to climatological reference forecast (based on all observations since 1965) and the observed inflow volumes for each respective year (black dots). Top panel: HBV volume forecasts resulting from current parameter set used in operational model (i.e. Curr) and optimal parameter set after calibration against inflow (i.e. HBV(*F1*)). Second panel: HYPE model volume forecasts resulting from optimal parameter set after calibration against inflow (i.e. HYPE(*F1, off*)). Third panel: relative volume error (ensemble mean) for HBV and HYPE forecasts shown in first two panels. Bottom panel: Continuous Ranked Probability Skill Score (CRPSS) for the HBV and HYPE model forecasts, using the climatological forecasts as reference. Spring flood volumes for each month are defined as the total reservoir inflow volumes between the 1<sup>st</sup> of each month until the 31 July (in day equivalent units, DE= 86,400 m<sup>3</sup>). Mean, min and max values are shown for each ensemble.**

Figures 3–5 show the CRPSS scores for individual forecast months and years for all combinations of model/calibration/initialization methods (refer to Table 3 in section 3.2.3). The latter provide further details about the response of the HBV and HYPE forecasts to the data assimilation:

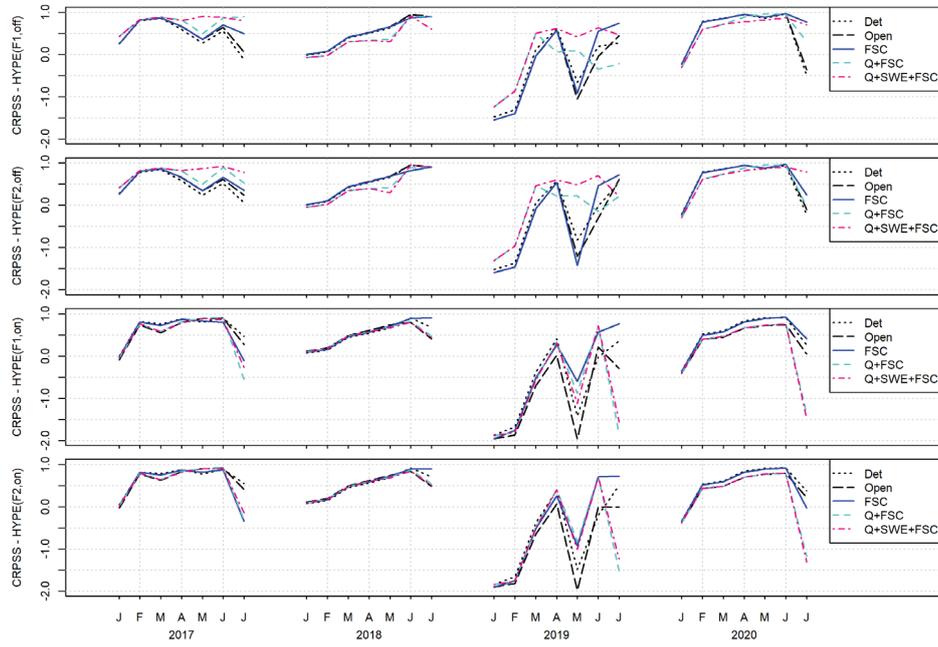
1. Impact of data assimilation is mainly seen from forecast month April–July, which makes sense since SWE observations were made in end of March, beginning of April, and there is little information content in the Q and FSC data before the onset of snow melt in April–May.
2. The *current* HBV model was mainly improved by updating the snow storage using the SWE observations. The auto-correction of the input data based on the Q data reduced the CRPSS in more cases than it was improved (Panel 1 in Figure 3).
3. The HBV(*F1*) model was improved by both the auto-correction of the input data and the SWE updating in all years except 2018, and was able to achieve CRPSS skill scores above 0 (better than the climatology) for all years and forecast months from March–July (Panel 2 in Figure 3).
4. The lower average CRPSS and larger improvement by SWE data assimilation found for the HBV(*F2*, *F4*) models compared to the HBV(*F1*) and HBV(*F3*) models, seems to be mainly a result of a low performance of the deterministic and Q-initialized models in 2019 (Panels 2–4 in Figure 3). In fact, the HBV(*F2*, *F4*) models seem to be consistently at a high CRPSS skill score during the other 3 years (2017, 2018 and 2019) (Panel 3 in Figure 3).
5. The HYPE (*F1*, *off*) model forecasts responded in a similar way to data assimilation of the input data and SWE assimilation as the HBV (*F1*) model: improved CRPSS in 2017 due to SWE assimilation, slightly reduced CRPSS in 2018 due to SWE assimilation, and largely improved CRPSS in 2019 thanks to the Q assimilation, and some minor improvements in 2020 (especially towards June/July) thanks to SWE assimilation (Panel 2 in Figure 3 and Panel 1 in Figure 4).
6. In general, the HYPE forecasts did not improve as much by the assimilation of the input data as it did for the HBV forecasts (Figures 3 and 4).
7. The HYPE model forecasts were usually less good for July compared with the HBV forecasts, but the ability to assimilate the FSC data did improve the HYPE model CRPSS values in July 2018, 2019 and 2020 (Figure 5).



**Figure 3: Continuously Ranked Probability Skill Score (CRPSS) for the HBV model spring forecasts with different calibration strategies (F1, F2-F4, and F3) and different data assimilation configurations: deterministic, Q (inflow auto-correction), SWE (SWE updating), and Q+SWE (inflow auto-correction and SWE updating). Calibration strategy is specified in the Y-axis label of each panel and its corresponding legend specifies the data assimilation strategy.**



**Figure 4: Continuously Ranked Probability Skill Score (CRPSS) for the HYPE model spring forecasts with different calibration strategies (F1-off, F1-on, F2-off, F2-off) and for the data-assimilation configurations in common with the HBV model experiments: deterministic, Q (inflow auto-correction), SWE (SWE updating), and Q+SWE (inflow auto-correction and SWE updating), as well as the open-loop ensemble reference (Open). Calibration strategy is specified in the Y-axis label of each panel and its corresponding legend specifies the data assimilation strategy.**



**Figure 5: Continuously Ranked Probability Skill Score (CRPSS) for the HYPE model spring forecasts with different calibration strategies (*F1-off*, *F1-on*, *F2-off*, *F2-on*) and for the data assimilation configurations involving fractional snow cover (deterministic, Open, FSC, Q+FSC, Q+SWE+FSC). Calibration strategy is specified in the Y-axis label of each panel and its corresponding legend specifies the data assimilation strategy.**

In Figure 6 we illustrate the functionality of the EnKF data assimilation as it is applied in the HYPE model. The figure shows the simulated inflow, snow water equivalent, and fractional snow cover with and without assimilation, as well as the temperature and precipitation for the January–July 2020. The red lines and symbols represent observations, black dashed line represent the deterministic model (without data assimilation), the blue line and shaded areas represent the EnKF model ensemble. The impact of the data assimilation is most visible in the SWE, where both the mean and the variance of the model ensemble were clearly changed at the snow observation date. There was also a small decrease in the simulated SWE due to the inflow data assimilation during the runoff event in January, and an accelerated snow melt due to the assimilation of inflow and fractional snow cover data in June and July. There were also clear impacts on both the simulated inflow and FSC during the same events. The impacts in terms of air temperature and precipitation are not so easy to see in these graphs, but there was a clear increase in air temperature during the snow melt period in June as a consequence of the inflow and/or fractional snow cover data assimilation.

A similar analysis was done for the HBV where different data-assimilation strategies were compared using the optimal parameter set found when calibrating only against inflow (i.e. HBV (F1), Figures 7–10). For each of the strategies, data assimilation was allowed from 20161201–20170701 by means on auto-correction of the input data and (or) manual correction of SWE in 2017-03-30 when data was available (i.e. SWE=763 mm). HBV model performances for the different data-assimilation strategies are found in Table 1. Manual correction of SWE together with auto-correction of the input data provided almost none improvement in model performance compared as if data assimilation was only done by auto-correction of the input data. There was some improvement to the deterministic model if manual correction of SWE was done, but auto-correction of the input data provided the largest improvement to the model.

**Table 1. HBV model performances based on different data-assimilation strategies for Överuman. Model performance correspond to the period 20170401–20170701.**

<b>Data assimilation approach</b>	<b>R2 (-)</b>	<b>Acc Diff (mm)</b>	<b>Rel Acc Diff (-)</b>	<b>Mean error (-)</b>	<b>r2log (-)</b>
Deterministic	0.81	19.77	0.039	0.40	0.69
SWE (manual)	0.86	-3.10	-0.005	0.34	0.70
Input (auto)	0.95	-0.96	-0.002	0.20	0.77
Input (auto) and SWE (manual)	0.95	-1.48	-0.002	0.19	0.77

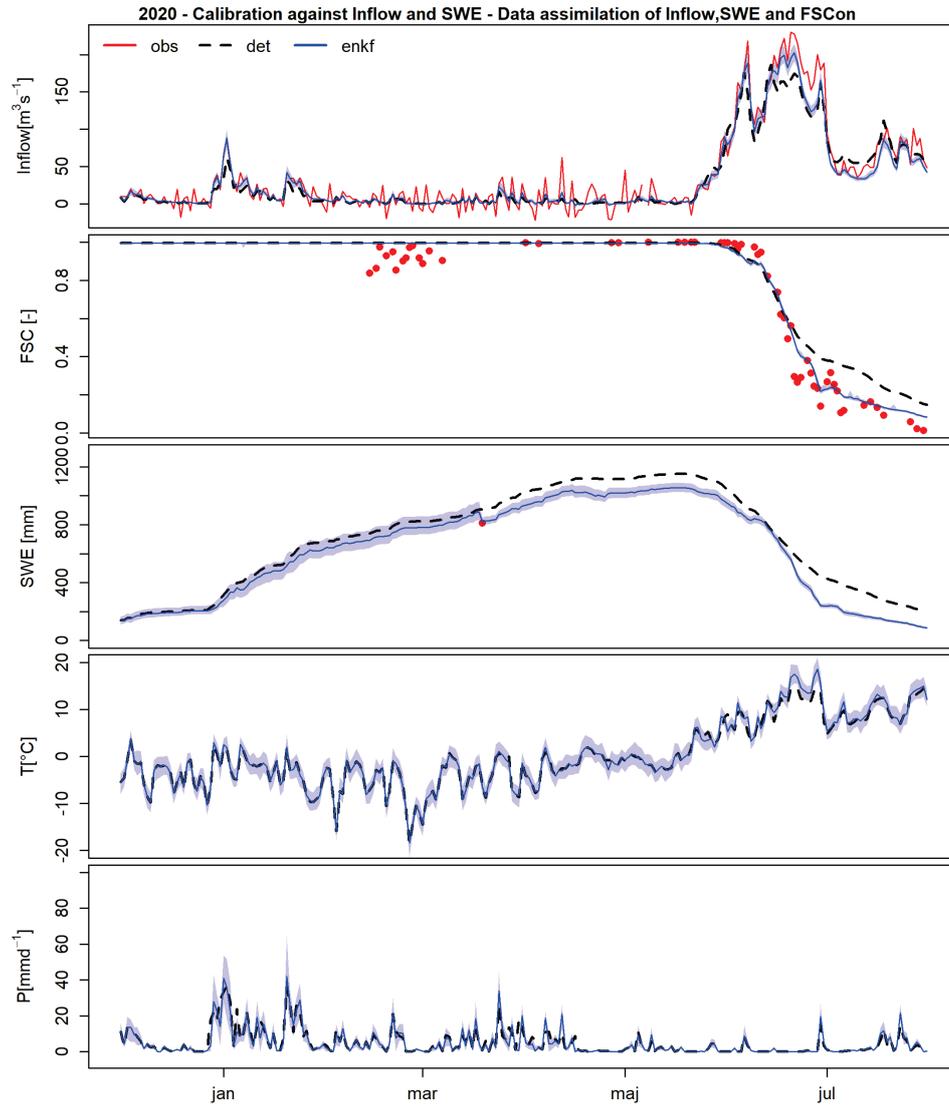
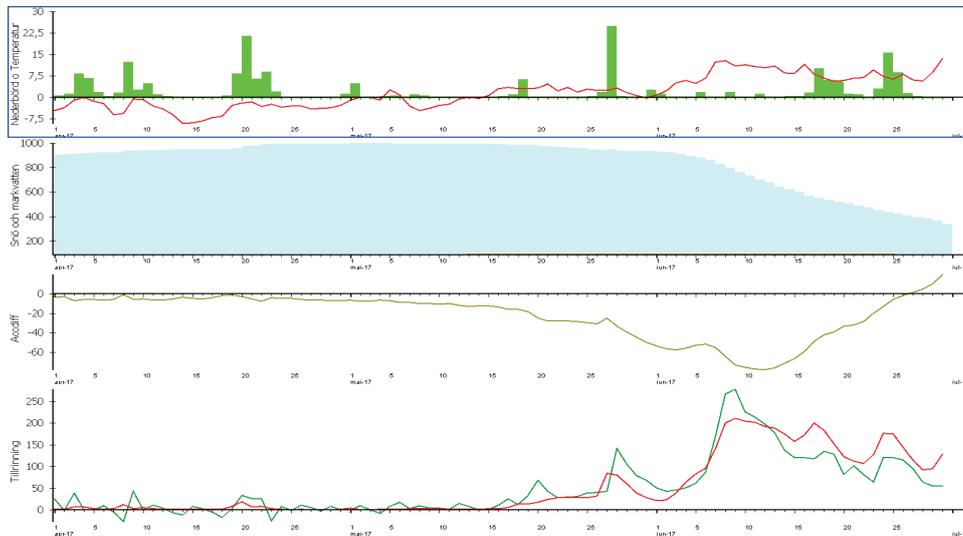
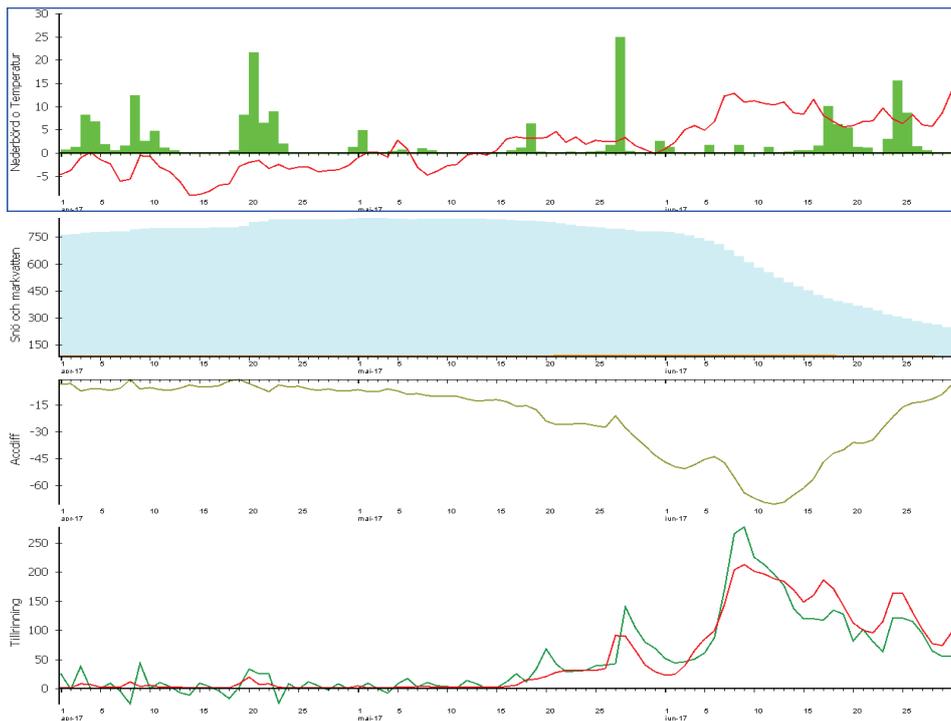


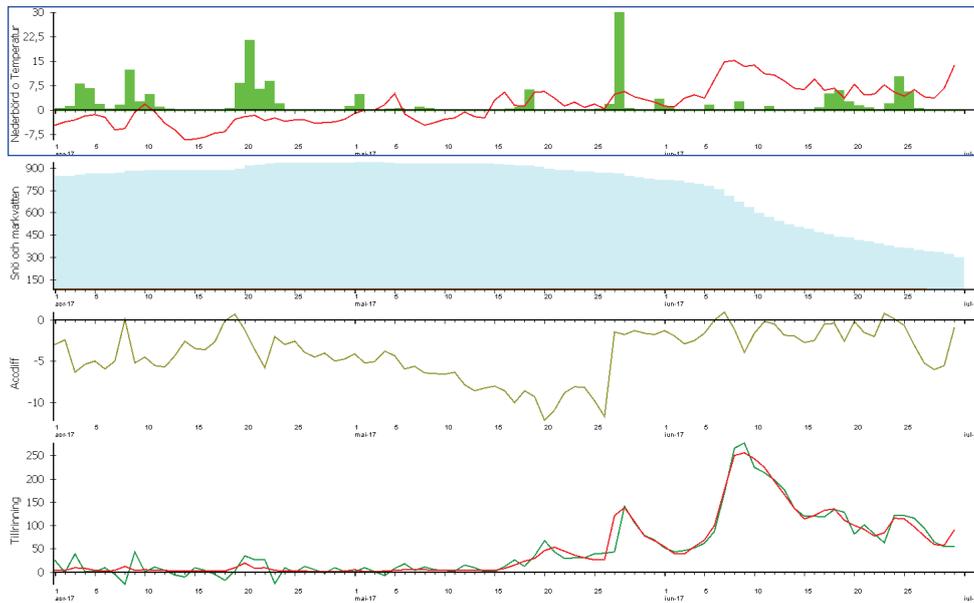
Figure 6: Impact of data assimilation on simulated inflow, snow water equivalent, fractional snow cover, and the meteorological conditions in the winter and spring 2020 in the HYPE model.



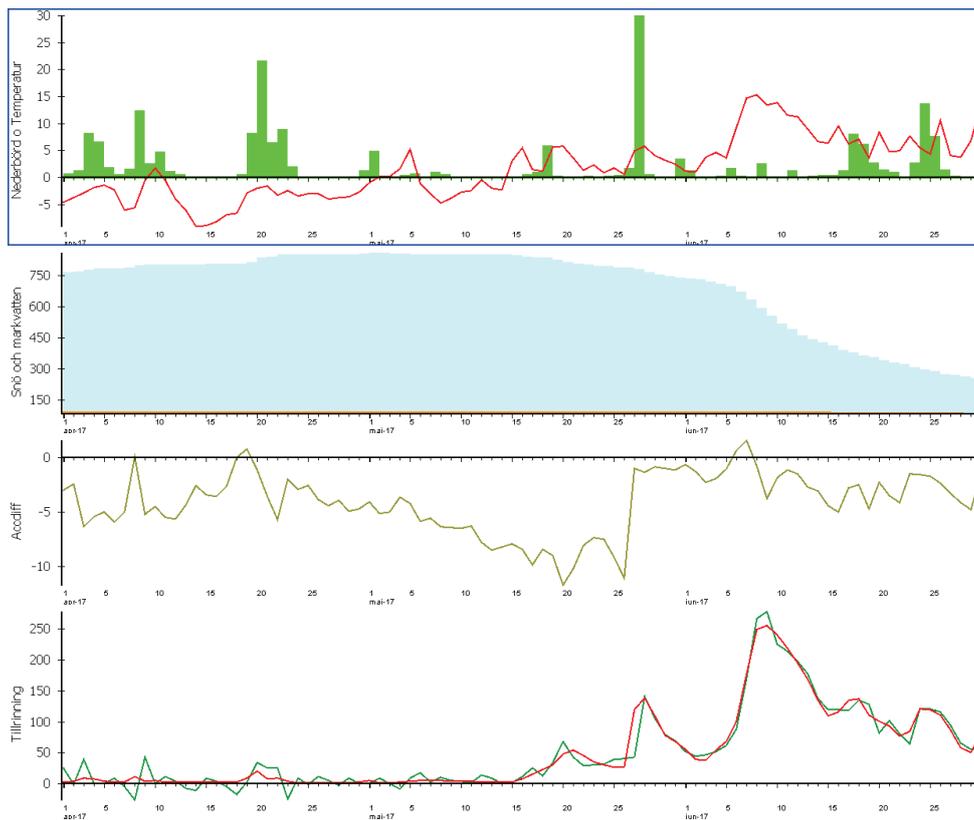
**Figure 7: HBV simulations from the deterministic model (without data assimilation) using the optimal parameter set obtained from calibration only against inflow.**



**Figure 8: HBV simulations after manual correction of SWE on 2017-03-30 using the optimal parameter set obtained from calibration only against inflow.**



**Figure 9:** HBV simulations after auto-correction of input data between 20161201–20170701 using the optimal parameter set obtained from calibration only against inflow.



**Figure 10:** HBV simulations after auto-correction of input data between 20161201–20170701 and manual correction of SWE on 2017-03-30 using the optimal parameter set obtained from calibration only against inflow.

# SPRING FLOOD PREDICTIONS WITH HBV AND HYPE

The focus here is to improve the precision of the hydropower industry's inflow forecasting models. The main purpose is to evaluate how different types of snow measurements can be integrated and used in hydrological modelling to better describe spring flows.

An important question in this field is whether a distributed model would perform better than a non- or semi-distributed model. Therefore, this project was developed to compare the two models HYPE and HBV, currently in use by SMHI. The sub-basin of lake Överuman was the study area for this experiment and the snow data available was snow water equivalent and fractional snow cover.

Assimilation of reservoir inflow, snow water equivalent, and fractional snow cover data individually or in combination during HBV and HYPE forecast model initialization generally improved the long-term spring flow forecast for all years and forecast periods from April-July. The combinations of assimilating both inflow and snow water equivalent data provided the highest forecast skill scores.

Energiforsk is the Swedish Energy Research Centre – an industrially owned body dedicated to meeting the common energy challenges faced by industries, authorities and society. Our vision is to be hub of Swedish energy research and our mission is to make the world of energy smarter!