Check for updates

Improving hydropower inflow forecasts by assimilating snow data

Jan Magnusson, Geir Nævdal, Felix Matt, John F. Burkhart
and Adam Winstral

ABSTRACT

Accurate long-term inflow forecasts are essential for optimal planning of hydropower production. In snow-rich regions, where spring snowmelt is often the largest reservoir of water, inflow forecasts may be improved by assimilating snow observations to achieve more accurate initial states for the hydrological models prior to the prognosis. In this study, we test whether an ensemble Kalman based approach is useful for this purpose for a mountainous catchment in Norway. For 15 years, annual snow observations near peak accumulation at three locations were assimilated into a distributed hydrological model. After the update, the model was run for a 4-month forecasting period with inflows compared to a base case scenario that omitted the snow observations. The assimilation framework improved the forecasts in several years, and in two of the years, the improvement was very large compared to the base case simulation. At the same time, the filter did not degrade the forecasts largely, indicating that though the updating might slightly degrade performance in some years, it maintains the potential for large improvements in others. Thus, the framework proposed here is a viable method for improving snow-related deficiencies in the initial states, which translates to better forecasts.

Key words | data assimilation, hydrological modelling, inflow forecasting, snow modelling

Jan Magnusson (corresponding author)
Norwegian Water Resources and Energy
  Directorate (NVE),
Oslo,
Norway
E-mail: jan.magnusson@gmail.com

Geir Nævdal
NORCE Norwegian Research Centre AS,
Bergen,
Norway

Felix Matt
Statkraft AS,
Oslo,
Norway

Felix Matt
John F. Burkhart
Department of Geosciences,
University of Oslo,
P.O. Box 1047, Blindern N-0316, Oslo,
Norway

Adam Winstral
WSL Institute for Snow and Avalanche Research
  SLF,
Davos,
Switzerland

INTRODUCTION

To limit the increase in air temperature due to ongoing global warming, we need to make the best possible use of low-carbon power energy sources to reduce greenhouse gas emissions. For hydropower, reliable long-term inflow forecasts help production planners to optimally utilize the available energy. In areas with significant snow accumulations, we hypothesize that it is possible to improve such inflow forecasts by incorporating snow measurements into the prediction system. In snow-rich regions of Norway,

hydropower companies routinely measure snow water equivalents (SWEs) at certain points in the catchment area near peak accumulation. Currently, the practice at Norwegian hydropower companies is to update forecasting models manually with these measurements. This procedure has major drawbacks such as limited reproducibility, unknown representativity of the measurements, lack of an objective and systematic approach for improving the forecasts and, finally, the method is also labor intensive. To overcome such shortcomings, we show how these measurements can be assimilated into a hydrologic forecasting model automatically using an objective method. Our main aim of this study is to test whether the

proposed data assimilation method can improve the predictive skill of the forecasting system for a lead time of weeks to months.

In many hydropower dominated regions, and particularly the Nordics, the largest proportion of inflow to hydropower reservoirs on an annual basis often comes during the spring snowmelt period. We can predict this peak inflow to the reservoirs using either parametric snow cover models (e.g. Ohmura 2001) or energy-balance snow models based on physical principles (e.g. Anderson 1976), in combination with a hydrological model describing the water flow through the catchment. In combination with seasonal weather forecasts or by using climatological records, both types of snow models can provide inflow forecasts for lead times of several months. However, the parametric models, also called temperature-index models, are often employed in operational settings over the more data demanding energy-balance models since they only rely on air temperature and precipitation data alone. Uncertainty in the forecasted inflows depends on errors in the model structure, meteorological forcing data and initial conditions (Kuczera 1983; Beven & Binley 1992; Vrugt *et al.* 2005). For short-term forecasts, the uncertainty in the initial state variables can be reduced by updating the model with observed discharge using, for example, the Kalman filter (Fjeld & Aam 1981). For long-term predictions covering the snowmelt and summer season, the assimilation of snow states themselves may be of greater importance than the direct assimilation of runoff. In this study, we aim to minimize the total uncertainty in long-term runoff predictions by reducing the error in the initial snow states. Foremost, the simulated snow states at peak accumulation can be prone to large errors as potential forcing data errors accumulate throughout the snow accumulation season (Gragne *et al.* 2015). This can eventually lead to large errors in the predicted inflows during spring snowmelt. Thus, reducing the errors in the initial snow conditions may yield an improved long-term inflow forecast.

A large range of methods exist for updating snow models using observations. With direct insertion, the simulated states are adjusted to exactly match the observed values at the same location (e.g. Liston *et al.* 1999; Fletcher *et al.* 2012). However, this method will likely produce poor results if the errors in the measurement are large, and direct insertion cannot be used to update the model at locations lacking observations. The latter problem can be circumvented by applying some interpolation method, for example, by deriving spatially distributed grids of correction factors for the model parameters (e.g. Liston & Hiemstra 2008). In order to take the observation uncertainty into account, we can choose one of several methods originating from Bayes' theorem. Kolberg *et al.* (2006) and Kolberg & Gottschalk (2006) presented a method for assimilating snow-covered area information into the snow routine of a runoff model based on Bayes' theorem. Another approach is to use so-called optimal interpolation, which has been used for improving continental snow maps by incorporating snow depth observations (e.g. Brown *et al.* 2003; Barnett *et al.* 2005) or remote sensing data (e.g. Liu *et al.* 2013). Such methods can be further improved by using approaches that make use of the information provided by the state-space model. Two examples of such methods that have been applied for snow models are the particle filter (e.g. Leisenring & Moradkhani 2011; Magnusson *et al.* 2017) and the ensemble Kalman filter (EnKF) (e.g. Andreadis & Lettenmaier 2006; Slater & Clark 2006; De Lannoy *et al.* 2012; Magnusson *et al.* 2014). With the latter method, we can update snow models at unobserved locations and also take uncertainties in the forcing data and snow observations into account.

The above-mentioned studies have shown that model states can efficiently be improved by assimilating either ground or satellite-based snow observations. However, whether the assimilation of snow data before spring snowmelt also improves long-term predictions of inflows to, for example, hydropower reservoirs is still an open question. Thus, the aim of this study is to assess the value of snow observations for improving inflow forecasts covering the spring and summer season. We present an assimilation framework using the EnKF in combination with a temperature-index snowmelt model coupled to a hydrological model. The performance of the system was benchmarked against a base case simulation that did not utilize the snow observations. We also assessed whether eventual improvements in inflow forecasts depended on factors such as the average snow amounts in the catchment and if the number of assimilated snow observations influenced the quality of the predicted inflows.
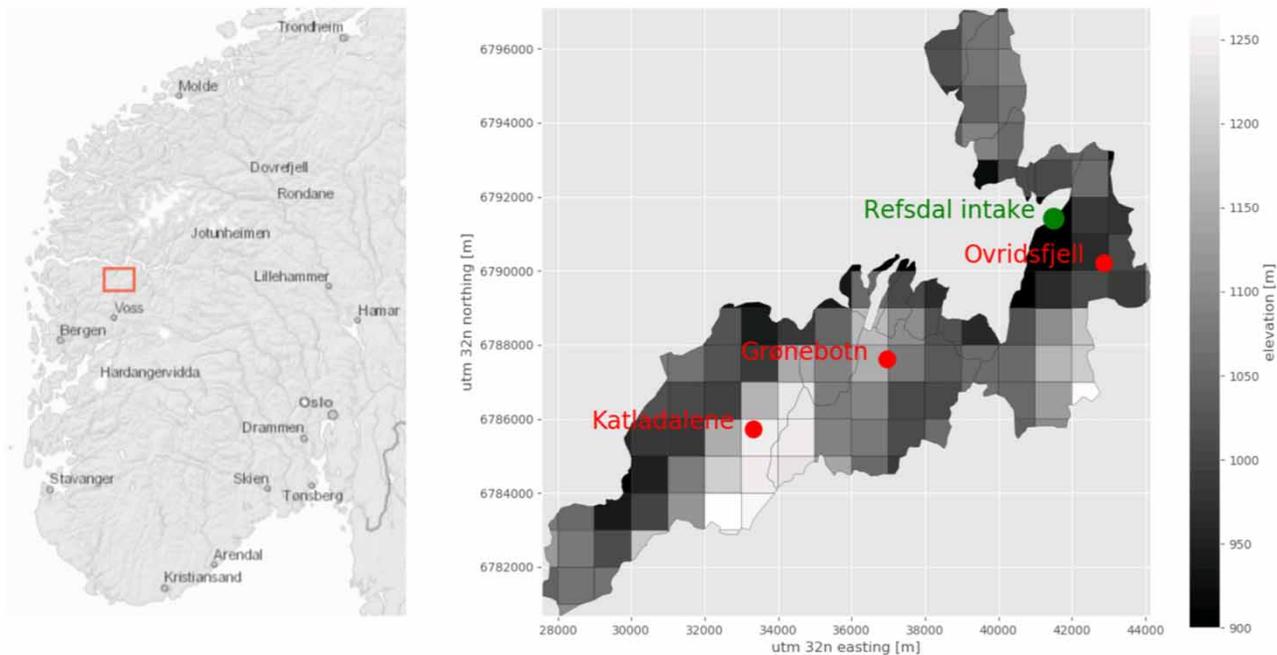
Figure 1 | The left panel shows the location of the Refsdal catchment in southern Norway, and the right panel displays the drainage basin for the catchment, including the three snow measurement locations (red dots) and the intake location of the Refsdal hydropower plant (green dot) where discharge is measured. The elevation of the grid cells used by the hydrological model is shown in the background of the right panel. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2020.025.

STUDY AREA AND DATA

Refsdal catchment

In this study, we used data covering the period from 2000 to 2015 for the Refsdal catchment (Figure 1). The watershed is located just south of the Sognefjord at the Norwegian west coast. The drainage basin covers an area of approximately 74 km^2 and altitudes range from 530 to 1,276 m.a.s.l. with an average elevation of 1,062 m.a.s.l. Hydropower reservoirs cover approximately 8% of the watershed area and forests cover a negligible part of the catchment (<1%). Thus, the catchment represents a typical mountainous landscape above the tree line at high latitudes.

For the study period, SWE has been measured at three locations (Katledalane – 1,280 m.a.s.l., Grønebotn – 1,060 m.a.s.l. and Ovridsfjell – 1,060 m.a.s.l.) throughout the watershed (see below for more information about those observations). Mean annual precipitation in the catchment equals approximately 2,200 mm, of which about 50% falls as snow.

Model forcing data

We used the seNorge Version 2 dataset of daily gridded average air temperature and precipitation for this study (Lussana & Tveito 2017). These gridded data have been generated for mainland Norway based on hundreds of observations statistically interpolated to a 1 km grid. The data records have been quality controlled, and detected errors have been corrected or removed using an automatic procedure. The precipitation data have, however, not been corrected for undercatch (Lussana et al. 2018), yet these data are inputs to operational hydrological models in Norway (Luijting et al. 2018) and have been used for research purposes (e.g. Huang et al. 2019). Thus, the precipitation grids likely underestimate the real-world precipitation, in some regions severely likely due to wind effects, low station coverage and poor representativity of the observation sites. For more details about the gridding procedure in seNorge for air temperature and precipitation and the quality of the grids, see Lussana et al. (2016) and Lussana et al. (2018), respectively.

Snow and inflow observations

Statkraft, a Norwegian hydropower company, conducts annual snow measurement campaigns at three locations inside the catchment. These measurement campaigns are typically conducted in early April. At each location, between 30 and 40 snow depth measurements are conducted in fixed intervals (10–25 m between each measurement, depending on topography) along straight transects at each site, and snow density is measured at a few points along the line. Each transect is positioned with the aim to provide a representative measurement for the area, and each campaign covering the three locations is undertaken within a couple of days. Per transect, an average SWE is calculated as follows:

$$\text{SWE} = \frac{1}{n}\sum_{i=1}^{n} h_i * \frac{1}{m}\sum_{i=1}^{m} \rho_i$$

where h is the measured snow depth in meters at n locations along the route and ρ is the bulk snow density in kg/m^3 at m sites along the same snow transect. This SWE value is used herein to update the model predictions.

In addition, naturalized catchment discharge based on hydropower production records from the Refsdal power plant has been made available to us by Statkraft.

METHODS

Model description

In this study, we used the open-source hydrological toolbox Shyft (https://gitlab.com/shyft-os/shyft) for the data assimilation experiments. This framework has been developed to provide operational inflow forecasts for the hydropower industry. Shyft uses distributed modeling concepts and is optimized for the efficient simulation of the hydrologic processes relevant to the aforementioned purpose. The user can choose between both conceptual and more physically based modelling approaches for depicting various hydrological processes. This framework has been used for analyzing how parameter uncertainty in

hydrological modelling influences reservoir inflow forecasts (Tweldebrahn *et al.* 2018).

In this study, the model was set up to run at a daily time step on a grid with a horizontal resolution of 1 km (see Figure 1). We simulated the snow cover development based on the methods applied in the Hydrologiska Byråns Vattenbalansavdeling (HBV) model (Lindström *et al.* 1997). This is a temperature-index based approach that also simulates and tracks the liquid water content of the snowpack as well as the snow distribution within each grid cell using a tiling approach. Potential evapotranspiration was computed using the Priestley–Taylor formula (Priestley & Taylor 1972). The subsurface response routine was based on the non-linear reservoir method presented by Kirchner (2009). In this study, routing of water between the individual grid cells was omitted since the catchment is small and we run the model on daily time steps.

For the simulations, we used the air temperature and precipitation data described above. A correction of precipitation amount is done through multiplying interpolated precipitation with a regional scaling parameter estimated during model calibration. This parameter should remove biases in the precipitation forcing that otherwise would severely degrade the simulations. Solar radiation, which is an input to the Priestley–Taylor algorithm and therefore affects the calculation of evapotranspiration, was set to a constant value since this part of the modelling chain has a minimal impact on the model performance during the snowmelt period for our study region. The model parameters were calibrated using the observed runoff for the period from 2009-9-1 to 2015-8-31. With the calibrated parameter values, the model shows a Nash–Sutcliffe efficiency (NSE) equal to 0.72 and a modified NSE of 0.65, for the whole study period. The modified NSE was computed using the climatology of observed discharge (monthly averages) as a benchmark instead of the average observed discharge following the methods outlined by Schaefli & Gupta (2007).

Data assimilation algorithm

For the data assimilation, we use a standard stochastic EnKF described briefly below. For further information about this filter, see e.g. Evensen (2009) and Vetra-Carvalho *et al.* (2018) and references therein. Let us begin by denoting the

state vector of the system as s_k, where k represents the time. (In our case, the state vector contains the states of all the grid cells of the catchment.) The filter consists of a forecast step and an analysis step. In the forecast step, we propagate our model, which we denote by f, forward in time giving $s_k = f(s_{k-1}, \delta_{k-1})$, where δ_{k-1} represents a stochastic forcing term. In our case, the state vector consists of SWE in each grid cell of the model and the forcing term is the weather inputs consisting of air temperature and precipitation at each grid cell. An ensemble of forward models is run, each having its own state vector and stochastic forcing. The forward model for ensemble member i is denoted $s_k^i = f(s_{k-1}^i, \delta_{k-1}^i)$. When the measurements, which in our case are observations of a subset of the states, become available at a generic time k, we perform the analysis step. In this study, the measurements represent average SWE at three grid cells within the simulation domain. The elevation differences between the observations and the grid cells are lower than 100 m, and biases between the measurements and simulations arising due to this small altitude discrepancy were not accounted for in the data assimilation scheme. In the analysis step, the covariance matrix P_k of the ensemble of state vectors $[s_k^1, s_k^2, \ldots, s_k^N]$ is required. Here, N denotes the size of the ensemble. The covariance matrix P_k is defined as follows:

$$P_k = \frac{1}{N-1} \sum_{i=1}^{N} (s_k^i - \overline{s_k})(s_k^i - \overline{s_k})^T$$

where

$$\overline{s_k} = \frac{1}{N} \sum_{i=1}^{N} s_k^i$$

is the ensemble mean at time k. Let R denote the covariance matrix representing the measurement uncertainty and let the observed quantities y_k be related to the states as $y_k = Hs_k$ for a matrix H at time k. (In our case H will be a matrix consisting only of zeros and ones and is formed such that it selects those states that are observed.) Then, the Kalman gain matrix is given as $K = P_k H^T (HP_k H^T + R)^{-1}$ where superscript T denotes the matrix transpose. Now, ensemble member i ($i = 1, \ldots, N$)

is updated by the formula $s_k^{i,a} = s_k^i + K(y_k - (Hs_k^i + \epsilon_k^i))$ where ϵ_k^i are samples from a Gaussian distribution with mean zero and covariance matrix R. The superscript a is used to distinguish the posterior ensemble members from the prior ensemble members. The matrix R is the covariance matrix of the measurement uncertainties. In this study, we used 50 ensemble members since increasing the number of ensemble members above this value did not reveal any relevant improvements, supporting the use of the fewer ensemble members to reduce computational costs.

Forcing ensemble generation

In this study, the stochastic forcing term in the filter algorithm outlined above was generated by applying perturbations to the temperature and precipitation data presented above. We assume that the errors in air temperature display much smoother variations in space than precipitation and have therefore chosen the approach below to construct the perturbations on the forcing data. For air temperature, the seNorge data were perturbed by adding normally distributed noise constant over the whole catchment (i.e. all the cells have the same perturbation). The noise had zero mean and a standard deviation of 2.0 °C and was correlated in time with a correlation coefficient equal to 0.9 between consecutive days. For precipitation, the forcing data were perturbed using multiplicative noise drawn from a lognormal distribution. This perturbation was produced in the following steps. First, we generated random correlated fields with a decorrelation length equal to 5 km using a fast Fourier transform algorithm. For each grid cell, the random numbers were normally distributed with zero mean and unit standard deviation. Second, we transformed this normally distributed noise to a lognormal distribution using the methods and parameters described in Magnusson *et al.* (2017). Finally, the precipitation grids were perturbed using this noise. We also introduced a correlation in time between the precipitation grids with a correlation coefficient of 0.5 between consecutive days. For more details about the choices of method and parameters for generating the noise on the forcing data, see Magnusson *et al.* (2014, 2017). This method could further be refined by increasing the perturbations during cold and windy conditions when the

precipitation gauges presumable measure precipitation with larger errors than during warm and calm conditions.

Description of experiments

Typically, long-term inflow forecasts are produced using either seasonal weather forecasts or some set of historical climatological data. In this study, the main aim is to assess whether improvements in the initial snow conditions also translates to more accurate long-term inflow forecasts during spring melt. Therefore, we isolate our analysis as much as possible on this part of the total uncertainty and try to minimize the effect of the remaining sources of errors.

As a base case, we first run the Shyft model using the seNorge weather data without any stochastic perturbations and for the whole period from 2000-9-1 to 2016-8-31. This simulation represents a typical reservoir inflow simulation for which historical data are available. We benchmark our data assimilation experiments against this simulation.

For the data assimilation experiments, we run the model with the stochastic forcings presented above until the snow observations become available during the first winter. Note that the snow measurements are typically performed in early April. At this point, we stop the model and update the SWE grids using the observations with the EnKF. We specify a

measurement uncertainty for the SWE observations using a standard deviation equal to 50 mm. After the update, which typically occurs in early April, we run the ensemble for a period of 120 days without applying stochastic noise to the forcings. This period is usually long enough that most of the snowmelts in the catchment and factors such as the initial snow distribution does not affect the results largely. We then compare the simulated inflow of the base case and the data assimilation run over this period against the observed inflow. Finally, we repeat this procedure for each winter over the whole simulation period resulting in 15 inflow forecasts of 120 days available for further analysis. The advantage of this approach is that we can isolate the effect of changing the SWE field based on the measurements and thereby evaluate the effect of our data assimilation algorithm directly.

RESULTS

Figure 2 shows simulated and observed SWE for three representative winters: one winter with low amounts of snow (2009/2010, average measured SWE equals 521 mm), one with medium amounts (2003/2004, 881 mm) and one with high amounts (2004/2005, 1505 mm). During 2003/2004
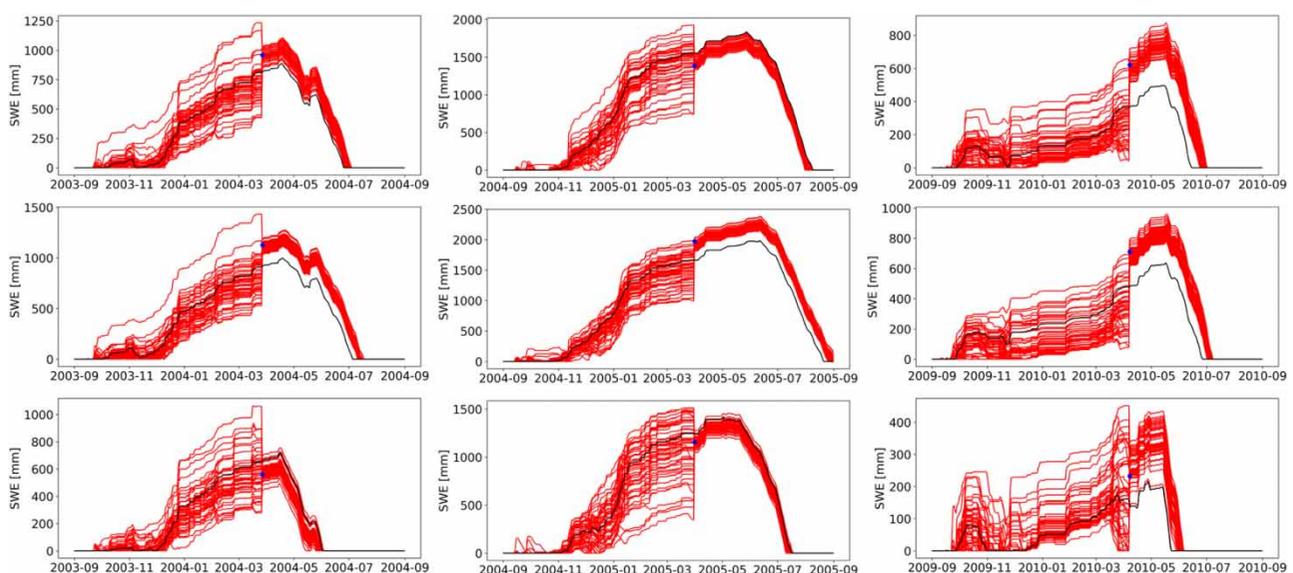

Figure 2 | Simulated and measured SWE for Grønebotn (top row), Katledalene (middle row) and Ovridsfjell (bottom row) for snow seasons 2003/2004 (left column), 2004/2005 (middle column) and 2009/2010 (right column). Ensemble members are shown in red, the base case in black and measurement as a blue dot. Note that the scales on the vertical axes differ between the panels. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2020.025.

and 2004/2005, the reference simulation matches the observations much better than in the drier winter (2009/2010), where the base run underestimates snow amounts at all three locations. The ensemble simulations show an increasing spread throughout the snow accumulation phase up to the point where the observations become available and encompass the observations in all years and locations displayed. The filter algorithm pushes the ensemble toward the observations and strongly reduces its spread, especially during the third winter. After the update, no perturbations have been added to the ensemble members to ensure better consistency with the base run during the 120-day long forecasting period and isolate the effect of the SWE update on the inflow forecasts. Therefore, the spread in the ensemble does not increase after the update.

As shown above, the filter algorithm adjusts the simulated SWE to better match the observations and improves the results compared to the base case (Figure 2). The question now arises whether this improvement in snow states also translates to better inflow forecasts. Figure 3 shows the forecasted cumulative discharge from the update time to the end of August for the same 3 years as displayed in Figure 2. In the beginning of the forecasting period, especially in 2005, both the base case and ensemble simulations tend to underestimate discharge, indicating that the model onsets snowmelt too late. Thus, the assimilation of snow measurements using the EnKF does not seem to improve the timing of simulated snowmelt. However, at the end of August, both the base run and ensemble simulations match the observed cumulative runoff rather well, in particular in 2005, even though of the large discrepancies in the start of the evaluation period. For 2004 and 2010, the

assimilation of SWE improves the predicted discharge throughout the summer period. The improvement is particularly large in 2010 when the base case performs worst. This is also the year in which the base case underestimated the observed snow amounts most (compare with Figure 2). Thus, it seems as if the assimilation of snow data improves the results in years where the snow simulations deviate largely from the observations.

Figure 4 shows the performance of the inflow forecasts for all years in the study period. The forecasting period covers 120 days following the point in time where we updated the ensemble using the snow observations (see illustration of this forecasting period in Figure 3 highlighted by gray vertical dashed lines). For each year, we computed the percentage bias in total inflow over the forecasting
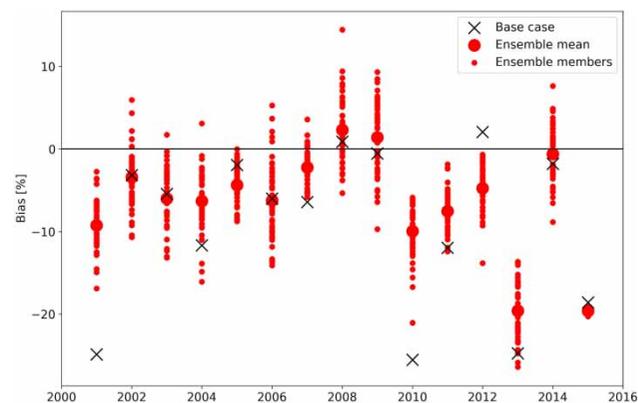

Figure 4 │ Error in total inflow to the reservoir during the forecasting period of 120 days for each year shown as the percentage of observed inflow for the base case (black cross), ensemble mean (larger red dot) and individual ensemble members (small red dots). The zero line is marked in black. A positive value means an overprediction, whereas negative values indicate an underestimation by the simulation. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2020.025.
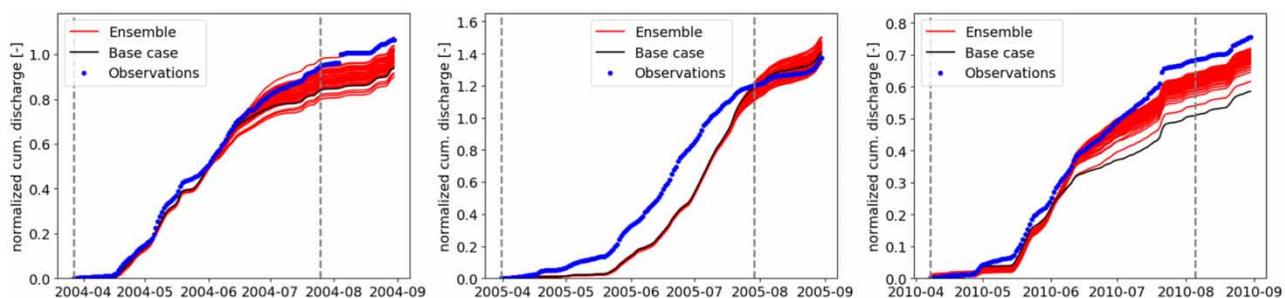

Figure 3 │ Forecasted cumulative discharge from the update time to the end of August for 2004 (left), 2005 (middle) and 2010 (right). A dashed line is shown at the point in time where the SWE observations were assimilated, and another one 120 days afterwards. Note that the presented values have been normalized by the average observed cumulative discharge for the forecasting period using data from all available years.

period for both the ensemble run and the base case using the observed inflow. The ensemble means show a lower absolute bias than the base case in 7 years (2001, 2004, 2007, 2010, 2011, 2013 and 2014). In two of those years, the improvement by updating the model is large (2001 and 2010). However, for 2015, the data assimilation algorithm does not improve the inflow forecast even though the error is large. The improvement for 2013 is also small compared to the magnitude of the error. For 3 years, the base case simulation shows a slightly better performance than the ensemble mean (2005, 2009 and 2012). The ensemble mean is within an error of approximately 10% for all but two of the seasons (2013 and 2015), whereas the base case run has more years with an error higher than 10% (2001, 2004, 2010, 2011, 2013 and 2015). Thus, the data assimilation scheme effectively removes large errors in the inflow forecast observed in the base case run.

For the three representative winters displayed in Figure 2, the greatest improvement in simulated SWE due to the assimilated data occurred in the year with the lowest snow amounts (2009/2010). In this winter, the base case underestimated SWE substantially at all three measurement locations, whereas the assimilation run showed a much better match with the observations (see right panels in Figure 2). This reduced error in SWE also contributed to a large improvement in the inflow forecast compared to the base case (see right panels in Figures 3 and 4). Figure 5
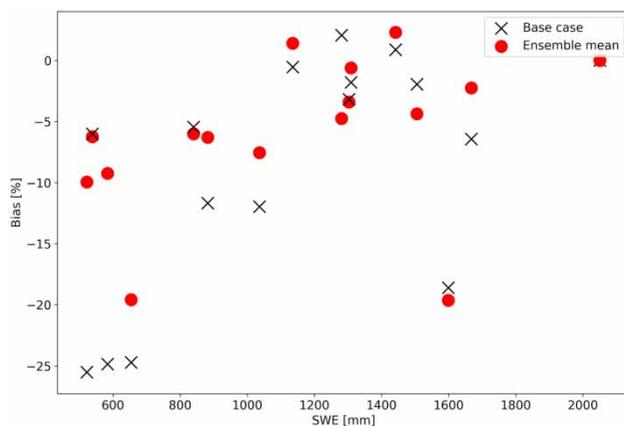
shows the performance in the inflow forecasts against the average measured SWE at the three locations both for the base case and the ensemble run. For high snow amounts (SWE greater than 1,100 mm), the assimilation of snow observations does not seem to improve the results much compared to the base case. For medium snow amounts (SWE between 700 and 1,100 mm), the ensemble mean shows a somewhat better performance than the base run in 2 out of 3 years. Notably, in this study we got the largest benefit of assimilating snow observations that occur for years with low amounts of snow (SWE below 700 mm). For these years, the filter algorithm improves the simulated inflows compared to the base case in 3 out of 4 years, and in two of those years, the improvement is very large. Thus, these results indicate that the assimilation of snow observations is more important for years with low than high snow amounts in the study catchment.

In this study, snow observations were available at three different locations throughout the watershed (see Figure 1). In an additional set of experiments, we assimilated snow observations using all possible combinations of available locations to test how sensitive the inflow forecasts are to the constellation of assimilated observations. The mean absolute error in forecasted inflow for all these simulations is shown in Table 1. The smallest error is achieved when assimilating data from both Grønbotn and Katladalene, whereas updating the simulations with data from Ovridsfjell alone gave the worst results. Using exclusively Ovridsfjell data, assimilated model performance was worse than the

Figure 5 | Bias in forecasted inflows against the mean of the measured SWE at the three locations. The bias in forecasted inflow is given as a percentage of the observed inflow to the Refsdal hydropower plant. The assimilation of SWE tends to improve the inflow forecast for years with low snow amounts more than for snow-rich years.

Table 1 | Mean absolute error (MAE) in forecasted inflow depending on the number of observations available for assimilation

Grønebotn	Katladalene	Ovridsfjell	MAE (%)
			10
X			7
	X		6
		X	15
X	X		6
X		X	8
	X	X	8
X	X	X	7

The crosses mark for which locations the snow observations have been used in the updating algorithm. Thus, the top row is identical to the base case, and in the last row, all measurements have been used.

base case, indicating that incorporating these SWE observations degrades inflow forecasts. However, when using data from all three locations the data from Ovridsfjell does not seem to affect the performance severely, indicating that the procedure is robust when using data from several locations. Nevertheless, with the rather pragmatic approach presented here, it can be determined which snow observations should be used for assimilation and which observations can be dropped or relocated to a better position in the measurement setup.

DISCUSSION

Even though the base case simulation matches the observed inflow well in many years during the study period, this simulation overall underestimates measured inflow, and in a few years severely (cf. Figure 4), even though the model includes a precipitation correction parameter. This underestimation is likely to a large degree, related to the precipitation data we have used as model forcing. For most regions in Norway, these data seem to underestimate precipitation (Lussana *et al.* 2018). Most important, the gauges are prone to undercatch, particularly of solid precipitation, leading to an underestimation of actual precipitation (e.g. Sevruk 1996; Wolff *et al.* 2015). Though the model includes a precipitation correction factor, the simulations underestimate observed inflow. Other factors influencing the quality of the precipitation forcing data are the sparse density of measurement stations and potential misclassifications of precipitation phase (i.e. that a rainfall event was classified as snowfall or vice versa). For some years (e.g. 2001, 2010, 2013 and 2015), the effects mentioned above seem to influence the base case simulation severely, and result in a large underestimation of the observed inflow.

In several of the years where the base case run provides poor results, the ensemble simulations improve the results (Figure 4). We attribute this improvement to the more accurate snow states at the beginning of the forecasting period since the setup of the two simulations are otherwise identical for this period. Note that during the forecast, both runs are driven with the same unperturbed input data. Since we update the model using snow observations made at peak accumulation, the filter likely reduces errors related to the

buildup phase of the snowpack, rather than processes related to ablation. Furthermore, we find that the assimilation of snow observations rarely degrades the inflow forecasts, and if so the difference with the base case is small. Thus, our setup of the filter algorithm seems successful and is a viable method for ensuring better inflow forecasts through more accurate initial states. The largest improvements in the forecasts are observed after winters with low amounts of snow (Figure 5). The updating using snow observations made near peak accumulation of the snow cover seems to reduce the errors in winters with the low amount of snow efficiently.

In addition to errors in the forcing data, poor results by the base case simulation may also be attributed to deficiencies in the snowmelt model. For some years, forecasted inflows occur with a delay compared to the observations (see, for example, 2005 in Figure 3). In fact, for all years with large snow accumulations (average measured SWE greater than 1,280 mm), the cumulative discharge forecasts lag behind observations at the onset of spring runoff. In this study, we have used a degree-day snowmelt model, including a bucket formulation for representing the routing of liquid water through the snowpack. This type of model does not capture all processes relevant for the onset of snowmelt and release of meltwater, and therefore introduces errors in the forecasts. However, in the case of seasonal inflow forecasting, the introduction of a more physically realistic model does not necessarily ensure better model performance. Such models require more variables as input, and those need to be of high quality in order to provide reliable results (e.g. Magnusson *et al.* 2015). In an operational setting, where the availability, representativeness and quality of the forcing data required by these physics-based solutions – whether provided by point observations or a weather forecasting model – add considerable uncertainties to the modeling chain (e.g. Raleigh *et al.* 2015). An alternative option for alleviating this problem might be to tune some of the parameters for the temperature-index snowmelt model using the EnKF during the ablation season by assimilating, for instance, the observed discharge data.

Utilizing the measurements from Grønebotn and/or Katladalene improved the forecasts, but the measurement from Ovridsfjell seemed to degrade the results. As there should not be any differences in how the measurements

are obtained, the explanation is likely either found in the data assimilation part of the workflow, or because single snow observations may not be representative for larger areas. Ovridsfjell is the most eastern of the three measurement locations, and a potential east-west gradient in the local climate conditions may cause this measurement to give limited information about snow conditions for the western part of the catchment. To avoid any unreasonable influences of Ovridsfjell for areas the measurement is not representative for, one could use localization in the EnKF (see Vetra-Carvalho *et al.* (2018) and references therein for more information about localization). With localization, the effect of a measurement is down-weighted in regions where it may not be representative but still influence the simulations in more similar regions. A second possible explanation might be that there is a height difference between the measurement location and the average height of the grid cell assigned to that location. However, it is not easy to adjust the observations to the grid cell altitude since elevation-dependent gradients in snow amounts can be complex (Grünewald *et al.* 2014). Finally, the measurement setup was not originally designed for assimilation in a distributed hydrological model that also represents the subgrid variability of the snow distribution using a tiling approach. It is still an open research question of how to best design a snow measurement campaign for updating such a model, and whether to use observed snow distributions or average amounts within each grid cell.

There are a number of decisions we have made while testing our data assimilation approach that will influence the results, much of which could be improved upon with further research. One crucial decision is related to how we generate the ensemble of SWE grids, using perturbations on the forcing grids that is required for the EnKF (or other ensemble-based data assimilation methods). Obviously, there are uncertainties associated with the weather forcings, and to reflect this, we have added perturbations to the historical weather data available from the seNorge archive. Ideally, our forcings should represent the uncertainty in these data, and for the time being, we are generating a rather *ad hoc* perturbation to the weather data. In the future, these *ad hoc* perturbations could be replaced by high-resolution ensembles given by the latest weather forecasting models. Such ensembles exhibit

physical consistency between the different variables due to the more realistic model used for generating the data and may also better reflect the uncertainty in the forcings. This may also improve the performance of the snow simulations, in particular when using an energy-balance approach, which is highly sensitive to biases in the inputs (Raleigh *et al.* 2015). However, the outputs from weather forecasting models likely need to be bias-corrected against ground observations to avoid large systematic errors before use in hydrological models. Furthermore, weather forecasting models also typically only provide data for the recent past making them unsuitable for the long-term analysis. Nevertheless, there exists a large potential for further testing and eventual improvement of the inflow forecasts by using better meteorological forecasts, snowmelt models and measurement setups.

CONCLUSIONS

In this study, we have assessed whether the assimilation of snow water equivalent (SWE) observations improves seasonal inflow forecasts for the snowmelt period. We assimilated snow observations, typically measured in the beginning of April, into a distributed conceptual hydrological model using an EnKF for 15 snowmelt seasons. The performance of the updating algorithm was tested by comparing simulated and observed inflows to a hydropower reservoir for 120-day forecast periods. The simulations with assimilated snow data were benchmarked against a base case that did not include the snow observations.

We find that the assimilation improves the seasonal inflow forecast for 7 years compared to the base case run, with a notable improvement for two of those years. For another 3 years, we observe a small decrease in model performance due to the assimilation. However, we do not see any larger degradations through the assimilation, indicating that though the updating might slightly degrade performance in some years, it maintains the potential for large improvements in others. In years where the assimilation does not improve inflow forecasts, this in fact may be related to factors other than inaccurate initial snow states, such as poor precipitation forcings during the forecasting period. In summary, the procedure tested here seems to work well

for reducing forecast errors that are related to deficiencies in the modelled snow states.

The performance of the data assimilation scheme depended on which snow observations were used for updating the model. Data from one of the measurement locations tended to degrade the inflow forecasts. However, the sampling scheme for the snow observations available in this study was not originally designated for updating a gridded model using an ensemble updating technique. For such a purpose, it is still not clear how to best design a sampling procedure for snow, and further research is warranted in this direction.

ACKNOWLEDGEMENTS

REFERENCES

Anderson, E. A. 1976 *A Point Energy and Mass Balance Model of a Snow Cover*. NOAA Technical Report NWS 19, Office of Hydrology, National Weather Service.

Andreadis, K. M. & Lettenmaier, D. P. 2006 Assimilating remotely sensed snow observations into a macroscale hydrology model. *Adv. Water Resour.* 29, 872–886. https://doi.org/10.1016/j.advwatres.2005.08.004.

Barnett, T. P., Adam, J. C. & Lettenmaier, D. P. 2005 Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature* 438, 303–309. https://doi.org/10.1038/nature04141.

Beven, K. J. & Binley, A. M. 1992 The future of distributed models: model calibration and uncertainty in prediction. *Hydrol. Processes* 6, 279–298. https://doi.org/10.1002/hyp.3360060305.

Brown, R. D., Brasnett, B. & Robinson, D. 2003 Gridded North American monthly snow depth and snow water equivalent for GCM evaluation. *Atmos.-Ocean* 41, 1–14. https://doi.org/10.3137/ao.410101.

De Lannoy, G. J. M., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E. C. & Pauwels, V. R. N. 2012 Multiscale assimilation of advanced microwave scanning radiometer – EOS snow water equivalent and moderate resolution imaging spectroradiometer snow cover fraction observations in northern Colorado: satellite-observed snow data assimilation. *Water Resour. Res.* 48. https://doi.org/10.1029/2011WR010588.

Evensen, G. 2009 The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Syst.* 29, 83–104. https://doi.org/10.1109/MCS.2009.932223.

Fjeld, M. & Aam, S. 1981 An implementation of estimation techniques to a hydrological model for prediction of runoff to a hydroelectric power station. *Model., Identif. Control* 2 (1), 37–56. https://doi.org/doi.10.4173/mic.1981.1.3.

Fletcher, S. J., Liston, G. E., Hiemstra, C. A. & Miller, S. D. 2012 Assimilating MODIS and AMSR-E snow observations in a snow evolution model. *J. Hydrometeorol.* 13, 1475–1492. https://doi.org/10.1175/JHM-D-11-082.1.

Gragne, A. S., Sharma, A., Mehrotra, R. & Alfredsen, K. 2015 Improving real-time inflow forecasting into hydropower reservoirs through a complementary modelling framework. *Hydrol. Earth Syst. Sci.* 19, 3695–3714. https://doi.org/10.5194/hess-19-3695-2015.

Grünewald, T., Bühler, Y. & Lehning, M. 2014 Elevation dependency of mountain snow depth. *Cryosphere* 8, 2381–2394. https://doi.org/10.5194/tc-8-2381-2014.

Huang, S., Eisner, S., Magnusson, J., Lussana, C., Yang, X. & Beldring, S. 2019 Improvements of the spatially distributed hydrological modelling using the HBV model at 1 km resolution for Norway. *J. Hydrol.* https://doi.org/10.1016/j.jhydrol.2019.03.051.

Kirchner, J. W. 2009 Catchments as simple dynamical systems: catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resour. Res.* 45, W02429. doi:10.1029/2008WR006912.

Kolberg, S. & Gottschalk, L. 2006 Updating of snow depletion curve with remote sensing data. *Hydrol. Processes* 20 (11). https://doi.org/10.1002/hyp.6060.

Kolberg, S., Rue, H. & Gottschalk, L. 2006 A Bayesian spatial assimilation scheme for snow coverage observations in a gridded snow model. *Hydrol. Earth Syst. Sci.* 10, 369–381. https://doi.org/10.5194/hess-10-369-2006.

Kuczera, G. 1983 Improved parameter inference in catchment models: 1. Evaluating parameter uncertainty. *Water Resour. Res.* 19 (5), 1151–1162. https://doi.org/10.1029/WR019i005p01151.

Leisenring, M. & Moradkhani, H. 2011 Snow water equivalent prediction using Bayesian data assimilation methods. *Stoch. Environ. Res. Risk Assess.* 25, 253–270. https://doi.org/10.1007/s00477-010-0445-5.

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

Liston, G. E. & Hiemstra, C. A. 2008 A simple data assimilation system for complex snow distributions (SnowAssim). *J. Hydrometeorol.* 9, 989–1004. https://doi.org/10.1175/2008JHM871.1.

Liston, G. E., Pielke, R. A. & Greene, E. M. 1999 Improving first-order snow-related deficiencies in a regional climate model. *J. Geophys. Res. Atmos.* **104**, 19559–19567. https://doi.org/10.1029/1999JD900055.

Liu, Y., Peters-Lidard, C. D., Kumar, S., Foster, J. L., Shaw, M., Tian, Y. & Fall, G. M. 2013 Assimilating satellite-based snow depth and snow cover products for improving snow predictions in Alaska. *Adv. Water Resour.* **54**, 208–227. https://doi.org/10.1016/j.advwatres.2013.02.005.

Luijting, H., Vikhamar-Schuler, D., Aspelien, T., Bakketun, Å. & Homleid, M. 2018 Forcing the SURFEX/Crocus snow model with combined hourly meteorological forecasts and gridded observations in southern Norway. *Cryosphere* **12**, 2123–2145. https://doi.org/10.5194/tc-12-2123-2018.

Lussana, C. & Tveito, O. E. 2017 seNorge2 dataset [Data set]. Zenodo. Available at: https://doi.org/10.5281/zenodo.845733.

Lussana, C., Tveito, O. E. & Uboldi, F. 2016 seNorge v2.0, Temperature (No. 14/2016). Norwegian Meteorological Institute.

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E. & Andersen, J. 2018 Senorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day. *Earth Syst. Sci. Data* **10**, 235–249. https://doi.org/10.5194/essd-10-235-2018.

Magnusson, J., Gustafsson, D., Hüsler, F. & Jonas, T. 2014 Assimilation of point SWE data into a distributed snow cover model comparing two contrasting methods. *Water Resour. Res.* **50**, 7816–7835. https://doi.org/10.1002/2014WR015302.

Magnusson, J., Wever, N., Essery, R., Helbig, N., Winstral, A. & Jonas, T. 2015 Evaluating snow models with varying process representations for hydrological applications: snow model evaluation. *Water Resour. Res.* **51**, 2707–2723. https://doi.org/10.1002/2014WR016498.

Magnusson, J., Winstral, A., Stordal, A. S., Essery, R. & Jonas, T. 2017 Improving physically based snow simulations by assimilating snow depths using the particle filter: snow assimilation with the particle filter. *Water Resour. Res.* **53**, 1125–1143. https://doi.org/10.1002/2016WR019092.

Ohmura, A. 2001 Physical basis for the temperature-based melt-index method. *J. Appl. Meteorol.* **40**, 753–761. https://doi.org/10.1175/1520-0450(2001)040<0753:PBFTTB > 2.0.CO;2.

Priestley, C. H. B. & Taylor, R. J. 1972 On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Weather Rev.* **100**, 81–92. https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH > 2.3.CO;2.

Raleigh, M. S., Lundquist, J. D. & Clark, M. P. 2015 Exploring the impact of forcing error characteristics on physically based snow simulations within a global sensitivity analysis framework. *Hydrol. Earth Syst. Sci.* **19**, 3153–3179. https://doi.org/10.5194/hess-19-3153-2015.

Schaefli, B. & Gupta, H. V. 2007 Do nash values have value? *Hydrol. Processes* **21**, 2075–2080. https://doi.org/10.1002/hyp.7072.

Sevruk, B. 1996 Adjustment of tipping-bucket precipitation gauge measurements. *Atmos. Res.* **42**, 237–246. https://doi.org/10.1016/0169-8095(95)00066-6.

Slater, A. G. & Clark, M. P. 2006 Snow data assimilation via an ensemble Kalman filter. *J. Hydrometeorol.* **7**, 478–493. https://doi.org/10.1175/JHM505.1.

Tweldebrahn, A. T., Burkhart, J. F. & Schuler, T. V. 2018 Parameter uncertainty analysis for an operational hydrological model using residual based and limits of acceptability approaches. *Hydrol. Earth Syst. Sci. Discuss.* 1–25. https://doi.org/10.5194/hess-2018-158.

Vetra-Carvalho, S., van Leeuwen, P. J., Nerger, L., Barth, A., Altaf, M. U., Brasseur, P., Kirchgessner, P. & Beckers, J.-M. 2018 State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems. *Tellus Dyn. Meteorol. Oceanogr.* **70**, 1445364. https://doi.org/10.1080/16000870.2018.1445364.

Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W. & Verstraten, J. M. 2005 Improved treatment of uncertainty in hydrologic modeling: combining the strengths of global optimization and data assimilation. *Water Resour. Res.* **41**, W01017. https://doi.org/10.1029/2004WR003059.

Wolff, M. A., Isaksen, K., Petersen-Øverleir, A., Ødemark, K., Reitan, T. & Brækkan, R. 2015 Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: results of a Norwegian field study. *Hydrol. Earth Syst. Sci.* **19**, 951–967. https://doi.org/10.5194/hess-19-951-2015.