# Can model-based data products replace gauge data as input to the hydrological model?

K. Sivasubramaniam, K. Alfredsen, T. Rinde and B. Sæther

## ABSTRACT

Hydrological models require accurate and representative meteorological inputs for better prediction of discharge and hence, the efficient management of water resources. Numerical weather prediction model-based reanalysis data products on the catchment scale are becoming available, and they could be an alternative input data for hydrological models. This study focuses on the applicability of a set of model-based data as input to hydrological models used in inflow predictions for operational hydropower production planning of three hydropower systems in middle Norway. First, the study compared the data products with gauge measurements. Then, Hydrologiska Byråns Vattenbalansavdelning (HBV) models of the three catchments were calibrated with three different meteorological datasets (model-based, gauge and observational gridded) separately using a Monte Carlo approach. It was found that the correlation between the model-based and gauged precipitation was highly variable among stations, and daily values showed a better correlation than hourly. The performance of model-based input data with daily timestep was nearly as good as the gauge or gridded data for the model calibration. Further, the annual simulated flow volume using the model-based data was satisfactory as similar to the gauge or gridded input data, which indicate that model-based data can be a potential data source for long-term operational hydropower production planning.

**Key words** | HBV model, hydropower production planning, inflow prediction, meteorological reanalysis, Monte Carlo calibration, numerical weather prediction (NWP) model

**K. Sivasubramaniam** (corresponding author)
**K. Alfredsen**
Department of Civil and Environmental Engineering,
Norwegian University of Science and Technology (NTNU),
7491 Trondheim,
Norway
E-mail: *kuganesan.sivasubramaniam@ntnu.no*

**K. Sivasubramaniam**
**T. Rinde**
Norconsult AS,
Postboks 626, 1303 Sandvika,
Norway

**B. Sæther**
NTE Energi AS,
Sjøfartsgt. 3, 7736 Steinkjer,
Norway

## INTRODUCTION

Today, precipitation-runoff models are employed as standard tools and routinely used for various hydrological applications (e.g. flood estimation, real-time flood forecasting, prediction of design flood and investigation of climate change and land use variability) (Wagener *et al.* 2004). Hydrological models combined with meteorological forecasts can provide a quantitative forecast of inflow to reservoirs and power plants, and it helps increase power production by reducing water spill and improving water management. Such models have been in operational use by hydropower companies in Norway since the 1970s, and they have proved to be cost-effective tools for hydropower operation and optimization (Killingtveit & Sælthun 1995). Calibration and updating of the states in a model are required before the model is used in an operational inflow forecast. The primary input data for precipitation-runoff models are typically time series of precipitation and air temperature with daily or hourly temporal resolution. Traditionally, *in situ* gauge observations are used as inputs for the models.

Hydrological models require accurate and representative meteorological inputs for better prediction and hence, the efficient management of water resources (Kirchner 2009; Beven 2012).



doi: 10.2166/nh.2020.076

Precipitation is an intermittent variable and various difficulties exist in obtaining quantitative precipitation precisely (Hwang *et al.* 2012). The measurements of precipitation using *in situ* gauges are subject to several error sources, such as wind-induced undercatch, wetting and evaporation losses (Førland *et al.* 1996; Taskinen & Söderholm 2016). The gauge measurement of solid precipitation (snow) in high latitudes and mountainous areas exhibits significant undercatch due to high wind conditions (Wolff *et al.* 2015). Further, the traditional *in situ* gauge observations represent point measurements and require a dense network of gauges to measure representative input on the catchment scale. However, in many areas, dense gauge networks are not common. In cases where existing sparse gauges do not capture the local precipitation distribution, the measured precipitation is not representative of the concerned catchment. Errors in the air temperature measurements are normally smaller, and the spatial variation of air temperature is also less; hence, air temperature observations from a station are generally more representative than precipitation (Ledesma & Futter 2017).

In some places (e.g. natural reserves, sanctuaries and remote mountainous areas), there are also restrictions and difficulties in operating *in situ* gauges. The Børgefjell national park in Norway which is located within the present study area is a typical example. Water draining from a 700 km$^2$ natural catchment is exploited for hydropower production. The power company is not able to install gauges within the nature reserve, and the hydrological model for inflow forecasting for Børgefjell is based on a single gauge located outside the area. Moreover, the operation and maintenance of precipitation gauges in remote mountainous areas incur considerable expenses.

Due to various challenges associated with the traditional approach of obtaining meteorological input data for hydrological models used in inflow predictions for operational hydropower production planning, hydropower companies in Norway seek alternative data sources for these purposes. Observational gridded datasets, remote sensing (weather radar and satellite) and numerical weather prediction (NWP)-based meteorological reanalysis data on the catchment scale can be potential alternative data sources to overcome challenges associated with traditional station data (Te Linde *et al.* 2008; Oke *et al.* 2009; Vu *et al.* 2012; Lauri *et al.* 2014; Ledesma & Futter 2017).

Observational gridded datasets are increasingly obtainable from the national and regional institutes (Haylock *et al.* 2008; Lussana *et al.* 2016; Lussana *et al.* 2018). Several studies have evaluated the observational gridded precipitation and air temperature datasets as model input compared to station data for medium- and large-scale river basins (Photiadou *et al.* 2011; Vaze *et al.* 2011; Essou *et al.* 2016a) and for small catchments (Ledesma & Futter 2017). Even though the gridded datasets have a continuous spatial coverage over the catchment and relatively fewer missing data compared to gauges, these datasets are generally derived from the available gauge measurements by spatial interpolation, and they have little additional information other than elevation (Essou *et al.* 2016a). Further, limitations in different interpolation techniques can also be a source of uncertainty (Vu *et al.* 2012; Lauri *et al.* 2014).

Precipitation measurements using remote sensing techniques (weather radar and satellite) are existing with high spatio-temporal resolution; however, these measurements of precipitation are indirect and subject to many sources of errors and uncertainties (Oke *et al.* 2009; Villarini & Krajewski 2010). Because of errors and uncertainties, the data from remote sensing techniques have not been widely used in operational hydrology so far (Berne & Krajewski 2013). Errors in the remote sensing are often corrected using ground-based gauge observations (Hasan *et al.* 2016; Sivasubramaniam *et al.* 2018); however, such corrections can only be possible in densely gauged regions.

In recent years, NWP model-based data products on the catchment scale with the increasing spatio-temporal resolution are becoming increasingly available as free and site-specific commercial products. The first guess forecasts from the NWP model are assimilated with the available past observations to make initial conditions for the next forecast. The same analysis for a fixed period produces meteorological reanalysis datasets with high spatio-temporal resolution (Talagrand 1997; Takahashi *et al.* 2010). Reanalysis datasets have been used in weather and climate studies (Takahashi *et al.* 2010) and used as atmospheric forcing data in hydrological models (Essou *et al.* 2016b). Compared with gridded data, the advantage of reanalysis data is that the dataset is updated regularly and available almost near real time (Essou *et al.* 2016b). However, errors and uncertainties related to reanalyses have not been understood well enough compared to those associated with gauge measurements (Parker 2016).

Previous studies have assessed the global and regional reanalysis datasets from different institutes and evaluated the use of them with hydrological models for runoff simulation (Te Linde *et al.* 2008; Lorenz & Kunstmann 2012; Vu *et al.* 2012; Lauri *et al.* 2014; Yang *et al.* 2014; Essou *et al.* 2016b; Roth & Lemann 2016). The focus of these studies was to use the reanalysis dataset as an alternative atmospheric forcing where the lack of gauge measurements exists. This study investigates the use of model-based data as input to hydrological models used in inflow predictions for operational hydropower production planning.

A typical inflow forecasting chain consists of the following components: (1) historical data to calibrate the models, (2) real-time data to update the current model states and (3) meteorological forecasts to generate inflow forecasts. Hydropower companies are involved in short-term and long-term operational planning. For short-term inflow forecasting, a calibrated and updated model is forced by 1–10 days of meteorological forecasts. Long-term predictions are normally run on average precipitation values taken from historical years to simulate a range of likely outcomes for the coming season or hydrological year. While short-term inflow forecasting is important for hydropower systems with low regulation capacity (runoff river schemes), long-term operational planning is required for well-regulated hydropower systems that consist of reservoirs with large volume relative to annual inflow, and snow-fed catchments where inflow prediction depends on snow storage in the catchments.

Hydropower companies usually buy the meteorological forecast data, used for daily inflow forecasting, from commercial weather service providers. The use of a gauge calibrated hydrological model with spatially defined prognosis data can also provide uncertainties in the predicted flow. The same providers now also produce meteorological reanalysis data as commercial products, and it is, therefore, of particular interest for hydropower companies to use them as a substitute for traditional gauge measurements since the model will be calibrated on data with the same spatial representation as the prognosis data.

The present study aims to answer two main research questions. First, can NWP model-based meteorological reanalysis datasets (precipitation and air temperature) replace traditional gauged precipitation and air temperature in the context of inflow predictions? Second, how do the model parameter and simulation uncertainty due to input data vary for the model-based data compared to the gauge and observational gridded data? To answer these questions, the study compares the time series of model-based data products with gauge observations at available gauge locations. Then, the study evaluates the performance of data products as an input to the hydrological model compared to the gauge and observational gridded datasets as an input. Further, the study analyses the uncertainty in the model parameters and the model response with the three forcing datasets.

## STUDY AREA AND DATA

### Study area

The model-based data were assessed over the Trøndelag region of central Norway. A Norwegian power company, Nord-Trøndelag Elektrisitetsverk (NTE) owns and operates more than 20 hydropower stations in this region, and its annual production is nearly 4,500 GWh. Three test catchments (Namsvatn, Follavatn and Tevla) with areas of 700, 200 and 350 km$^2$, respectively, are used in the setup of the Hydrologiska Byråns Vattenbalansavdelning (HBV) model in order to evaluate the performance of model-based data as an input. These are the major catchments in the NTE production system. The three catchments are shown in Figure 1, and their basic characteristics are presented in Table 1.

Based on the climatology for the period from 1961 to 2017, the mean annual precipitation in the study region is 2,000–4,000 mm along the coast and 750–2,000 mm inland. The annual mean temperature is in the range of 2–8 °C along the coast, and it is –4 to 2 °C in the inland mountainous areas (http://www.senorge.no/).

### Data

Gauge, observational gridded and model-based precipitation and air temperature data and river flow records from January 2010 to December 2016 were used in the present study.

NTE operates its own meteorological stations and uses the data (precipitation and air temperature) from them for

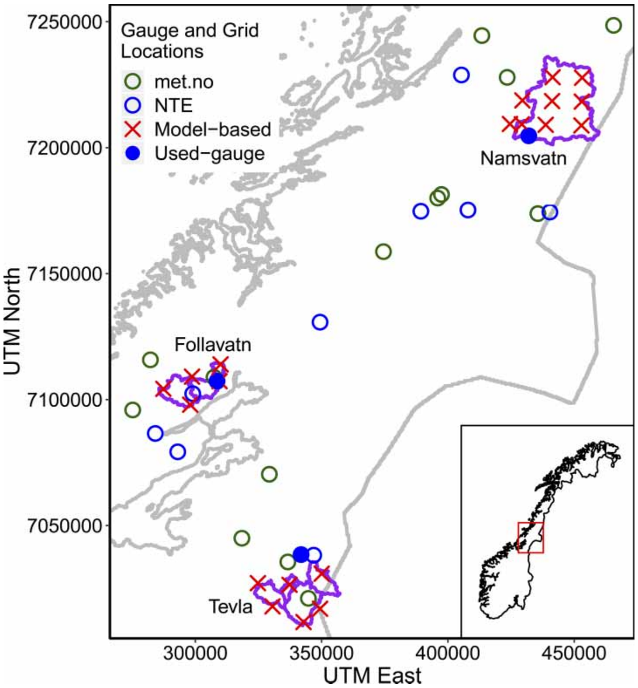


**Figure 1** | The weather stations operated by met.no (green circles), NTE (blue circles) and grid points of model-based data and three catchments (purple polygons) used in the study. The weather station (NTE), used for hydrological modelling of each catchment, is marked with a filled blue circle. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2020.076.

**Table 1** | Characteristics of the study catchments (source – http://nevina.nve.no/)

| Description | Namsvatn | Follavatn | Tevla |
|---|---|---|---|
| Area (km$^2$) | 701.5 | 202.7 | 345.9 |
| Elevation range (m.a.s.l) | 439–1,675 | 180–660 | 110–343 |
| River slope (m km$^{-1}$) | 9.5 | 12.5 | 24.2 |
| Forest (%) | 19.0 | 37.2 | 42.3 |
| Wetland (%) | 6.0 | 14.1 | 24.8 |
| Agriculture (%) | 0.1 | 0.9 | 0.7 |
| Bare mountain (%) | 62.2 | 34.3 | 27.3 |
| Lake (%) | 12.5 | 13.5 | 4.7 |
| Glacier (%) | 0.2 | 0 | 0 |
| Urban (%) | 0 | 0 | 0.2 |

inflow simulations. Within the study area, there are 12 NTE stations with the available hourly observations for the study (Figure 1). Besides, observations from the 14 weather stations operated by the Norwegian Meteorological Institute (met.no) were also used for the comparison with model-based data. Out of the 14 met.no stations, two of them are available with hourly precipitation and four with hourly temperature, and the rest are with daily observations.

The hourly time series of inflow data for the three catchments were obtained from NTE. The discharge values are back calculated. NTE calculated the net outflow draining from the catchments using the measurements of water level in the reservoirs and intake and transfer of water from and to the catchments. It can be noted that NTE has used these flow data in its operational HBV model.

StormGeo (http://www.stormgeo.com) commercially distributes meteorological forecasts to hydropower companies including NTE in Norway. Currently, StormGeo generates and distributes NWP model-based meteorological reanalysis using the MESAN (Mesoscale analysis model) (Häggmark *et al.* 2000) from the Swedish Meteorological and Hydrological Institute (SMHI) to its customers as site-specific end-user data products. MESAN assimilates NWP with ground observations (gauge and weather radar) to generate a meteorological reanalysis dataset. Here, NWP from the High-Resolution Limited Area Model (HIRLAM) were used. HIRLAM uses NWP from the European Centre for Medium-Range Weather Forecasts (ECMWF) as boundary conditions. The MESAN analysis model provides a dataset with 11 km $\times$ 11 km spatial resolution.

From MESAN analysis, StormGeo provided model-based hourly precipitation and air temperature data on the representative grid locations that spatially covers each of the catchments (Figure 1). It can be noted that StormGeo has distributed daily meteorological forecasts to NTE at these grid locations for operational model runs for the study catchments. In addition, StormGeo derived hourly precipitation and air temperature at the nearest model grids to the 26 meteorological stations (Figure 1) in order to compare the time series of model-based data with gauge observations in this study.

The Norwegian Meteorological Institute spatially interpolated the past observed precipitation and air temperature records from meteorological stations to develop the daily gridded (1 km $\times$ 1 km) precipitation (Lussana *et al.* 2018) and hourly and daily gridded air temperature (Lussana *et al.* 2016) datasets covering Norway. These datasets are freely available to the public through met.no's thredds server (http://thredds.met.no/thredds/catalog.html). The gridded precipitation and air temperature were downloaded for each study catchment. Hereafter, the NWP model-based reanalysis dataset from StormGeo is referred to as 'model-based' and observational gridded data as 'gridded' throughout the study.

## METHODS

### Data comparison

At each gauge location, the time series of model-based hourly precipitation and air temperature data were compared with the available hourly gauged observations. In addition, model-based hourly datasets were aggregated to daily and then compared with daily gauged data.

### HBV model

The HBV precipitation-runoff model is a semi-distributed conceptual model. A detailed description of the HBV model structure can be found in the literature (Bergström 1976; Bergström 1992; Killingtveit & Sælthun 1995; Sælthun 1996). The HBV model has been widely used in the Nordic region and other parts of the world for various hydrological studies (Steele-Dunne *et al.* 2008; Te Linde *et al.* 2008; Lawrence & Haddeland 2011). Most of the hydropower companies in Norway use a version of the HBV model for inflow forecasting.

In this study, PINEHBV (Rinde 1999), a variant of HBV, was used. The PINEHBV is in a structure similar to the model used by NTE. The model consists of four main storage components such as snow and soil moisture routines and two linear response tanks, upper and lower. The upper and lower zones generate the surface runoff and base flow, respectively. An illustration of the structure of the HBV model is added to Supplementary Figure S1 in Supplementary Materials. In the snow routine, the catchment is divided into ten elevation zones in order to account for the elevation-dependent variability in the type and amount of precipitation and snow storage. Further, among the ten zones, the lowest zones below the forest line based on the topography are defined as forested, and the remaining zones are non-forested. Determining the type of precipitation (snow or rain) and calculation of snowmelt and snow accumulation in each of the ten zones are the main processes in this component. The processes in the rest of the storage components are lumped at the catchment scale. Input to the PINEHBV model is the time series (daily or hourly) of precipitation and air temperature and monthly average potential evaporation.

### Performance evaluation of datasets

HBV uses a single input series of areal precipitation and temperature. We spatially averaged the model-based data from StormGeo grid points (Figure 1), and areal precipitation was estimated for each catchment. The operational HBV model at NTE uses observations from a single gauge for each catchment, and the same gauges (Figure 1) were used in this study and considered as the reference model. In addition, a spatial average of daily observational gridded precipitation of a regular grid (1 km × 1 km) was computed and used as a third input alternative.

For all three catchments, the same monthly average potential evaporation values were used, as shown in Table 2.

For each of three catchments, the HBV model was calibrated separately using gauge, observational gridded and model-based precipitation, and air temperature datasets. Since the study catchments are snow-fed, the start of the simulation was set to September to ensure no initial snow storage. Four years of data from September 2010 to August 2014 was used for the model calibration, and the model performance was evaluated for the three forcing datasets using a two-year verification period (September 2014–August 2016).

The Nash–Sutcliffe efficiency (NSE) (Nash & Sutcliffe 1970) was used as an objective performance criterion to evaluate model performance. The NSE is the most commonly used performance measure in hydrology (Essou

**Table 2** | Monthly values of daily potential evaporation

| Month | January | February | March | April | May | June | July | August | September | October | November | December |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Daily potential evaporation (mm/day) | 0.1 | 0.2 | 0.7 | 1.0 | 2.3 | 3.5 | 3.5 | 2.3 | 1.0 | 0.7 | 0.2 | 0.1 |

**Table 3** | Parameter ranges used in MC calibration

| Parameter | Description | Unit | Minimum | Maximum |
|---|---|---|---|---|
| RCORR | Precipitation correction factor, rainfall | – | 0.3 | 1.6 |
| SCORR | Precipitation correction factor, snowfall | – | 0.3 | 2.5 |
| TX | Threshold temperature for rain/snow | °C | −4.0 | 8.0 |
| CX | Melt index (degree day factor) | mm/°C day | 0.3 | 25.0 |
| CXN | Melt index – forest zones | mm/°C day | 0.3 | 25.0 |
| TS | Threshold temperature for melt/freeze | °C | −4.0 | 8.0 |
| TSN | Threshold temperature for melt/freeze – forest zones | °C | −4.0 | 8.0 |
| FC | Field capacity | mm | 5.0 | 1,500.0 |
| BETA | Relative contribution to upper zone from soil storage | – | 0.1 | 12.0 |
| FCDEL | Threshold value for potential evapotranspiration in soil moisture | – | 0.1 | 1.0 |
| KUZ2 | Upper recession coefficient, upper zone | mm/day | 0.1 | 5.0 |
| KUZ1 | Middle recession coefficient, upper zone | mm/day | 0.1 | 1.0 |
| KUZ | Lower recession coefficient, upper zone | mm/day | 0.01 | 0.6 |
| KLZ | Recession coefficient, lower zone | mm/day | 0.001 | 0.15 |
| UZ2 | Upper threshold, upper zone | mm | 5.0 | 500.0 |
| UZ1 | Lower threshold, upper zone | mm | 5.0 | 100.0 |
| PERC | Percolation constant upper to lower zone | mm | 0.0 | 5.0 |

*et al.* 2016a). In addition, accumulated flow difference (AccDiff) was used as an additional measure.

It is often shown that many different parameter sets can give similar good NSE (Beven & Binley 1992), and it is not given that the parameter set with the best NSE during the calibration provides good performance outside the calibration period (Seibert 1997). Therefore, a Monte Carlo (MC) approach of the model calibration was used to investigate how the uncertainty of the HBV model parameters varies for the three forcing datasets. The advantage of the MC is that the resulting parameter sets are not only a basis for investigating the model parameter uncertainty but also the simulated flow, and other model responses can be provided as a range instead of a single value (Steele-Dunne *et al.* 2008). Using the MC approach, Ledesma & Futter (2017) assessed the observational gridded data product compared to gauge measurements as the hydrological model input. Steele-Dunne *et al.* (2008) applied the MC method to generate an ensemble of simulated flows to assess the impacts of climate change on hydrology.

Each of the free parameters (17 parameters) in the PINEHBV model was given a range of reasonable values, as suggested in earlier studies (Killingtveit & Sælthun 1995; Sælthun 1996; Rinde 1999) and shown in Table 3. An MC model calibration with uniform sampling (Seibert 1997; Seibert 2003; Steele-Dunne *et al.* 2008; Ledesma & Futter 2017) was undertaken to generate an ensemble of 100,000 parameter sets for each of the three catchments using the three forcing datasets separately. From the 100,000 parameter sets, the best 100 parameter sets with the highest NSE were chosen, and then, from those 100 parameter sets, the best 50 parameter sets which also give the highest NSE during the verification period were finally selected. An ensemble of the 50 simulated model responses with the 50 best parameter sets were used for the analysis.

## RESULTS

### Data comparison

For comparing model-based precipitation and temperature with gauge observations, a linear regression analysis was carried out. Pearson correlation coefficient between model-based and gauge data was calculated at each gauge location. Figure 2 shows the box plot of the estimated

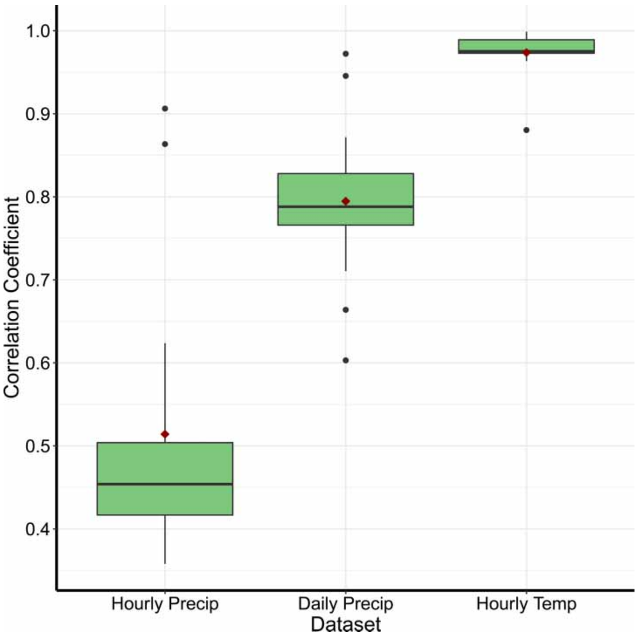


**Figure 2** | Box plot of the correlation coefficient between model-based and gauge precipitation and air temperature data with an hourly and daily resolution, estimated at gauge locations. The values outside 1.5 × IQR are represented by the whiskers.

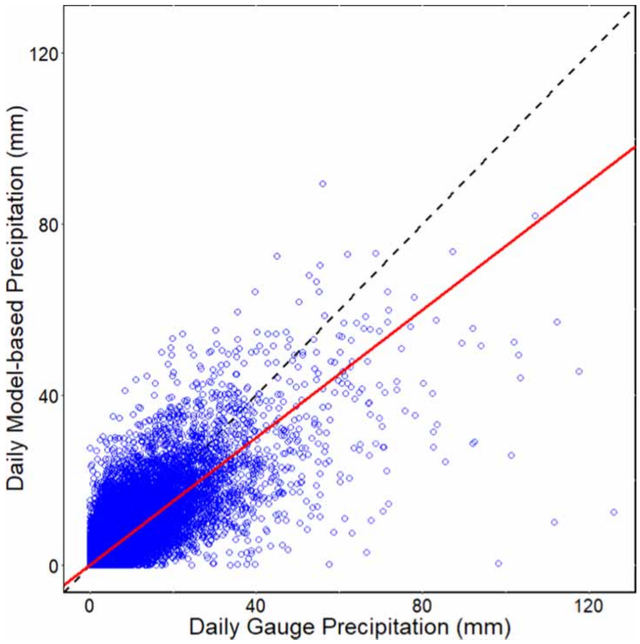


**Figure 3** | Scatter plot of model-based and gauge precipitation data, pooled from all gauge locations. The dashed line denotes the perfect fit 45-degree line, and the red solid line shows the regression.

correlation coefficient between the model-based and gauge datasets.

Looking at Figure 2, the model-based hourly precipitation shows a poor correlation with the gauge observations. For the two met.no stations available with hourly measurements, the correlation is relatively high (shown as outliers in Figure 2). For all NTE stations, the hourly precipitation data show a poor correlation. However, daily precipitation data show a reasonably good correlation with a few exceptions. For all gauge locations, the hourly model-based temperature correlated well with the gauge measurements.

We prepared scatterplots and compared the model-based and gauged datasets at each gauge location. Figure 3 shows a single scatterplot of all data pooled together. Looking at Figure 3, it is particularly seen that the model-based data in most cases underestimate high-intensity daily precipitation events observed by gauges.

To investigate how the precipitation volume of the three forcing datasets vary and how this variation influences the model performance, we compared the accumulated annual precipitation of the three forcing datasets over the three catchments. Figure 4 shows that the model-based and

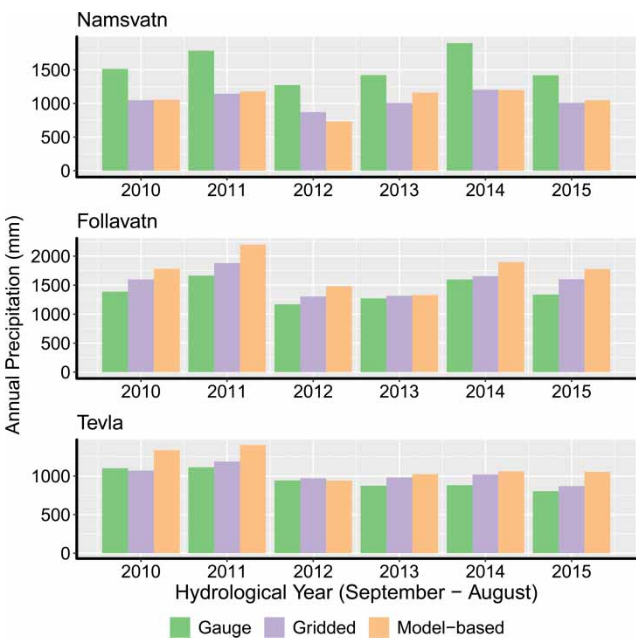


**Figure 4** | Annual precipitation from the gauge, gridded and model-based precipitation input data for the three catchments during the calibration (September 2010–August 2014) and verification period (September 2014–August 2016).

gridded precipitation are similar in volume for Namsvatn. It is also seen that for the Namsvatn catchment, the model-based and gridded precipitation are lower than the gauge precipitation, while for Follavatn and Tevla catchment, they are higher than the gauge value, and the model-based yielded the highest annual volume for almost all years for these two catchments.

## Performance in simulating the hydrological response

The performance of flow simulation (NSEs) of the three forcing datasets (gauge, gridded and model-based) is shown in Figure 5. The NSE value above 0.6 is generally considered as an acceptable model by hydrologists (Essou *et al.* 2016b). As presented in Figure 5, the NSEs for the 50 optimum parameter sets during the calibration and verification period for all three forcing datasets were mostly above 0.6 in this study except for model-based data on the Follavatn catchment (NSEs during the verification period in the range of 0.52–0.66). Looking at Figure 5, the performance of gauge and gridded during the calibration period is higher than model-based for Namsvatn and Follavatn, while the performance of model-based was superior to that of the gauge and gridded for the Tevla catchment. The best NSEs for the gridded dataset during the calibration period for Namsvatn resulted in a relatively lower NSE during the verification period. It was found that a parameter set with the best NSE during the calibration did not give the best NSE during the verification; a parameter set slightly lower NSE than the best one gave a better NSE during the verification.

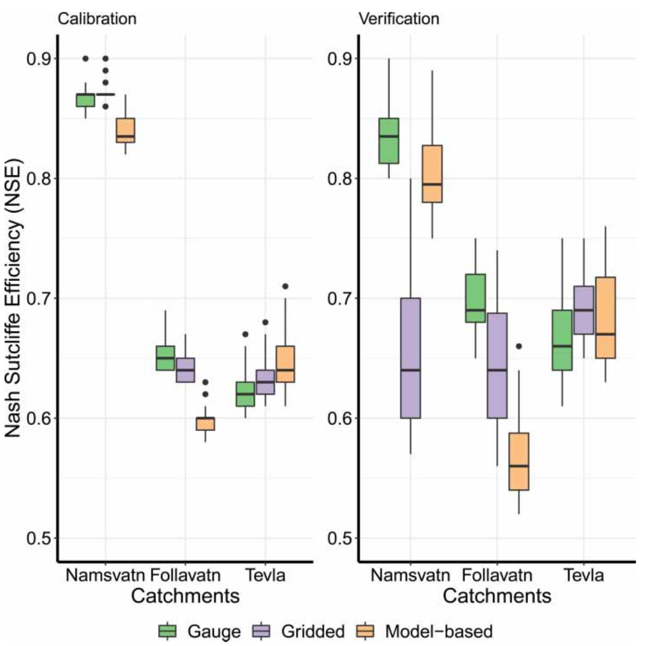


**Figure 5** | Box plot summary of the NSEs of 50 optimum parameter sets for the three forcing datasets (gauge, gridded and model-based) for the three catchments during the calibration and verification period.

While Figure 5 presents a summary of NSEs for daily flow, Table 4 presents the median values of NSEs, estimated from the simulated flows with daily and weekly timescale using the 50 optimum parameter sets. Looking at Table 4, the performance (NSE) on a weekly temporal scale is higher than the daily, and it is over 0.67 for all three datasets. Like daily timescale, the performance of the three forcing datasets showed a nearly similar variation on the weekly timescale.

An ensemble of 50 simulated flows using the 50 optimum parameter sets for each of the forcing datasets for the three catchments is shown in Figure 6. Here, the verification period is presented to illustrate the responses for the three forcing datasets; a plot for the calibration period (Supplementary Figure S2) is added to the supplementary material. In Figure 6, three colour bands represent the ensemble of simulated flows for the three forcing datasets. If blue or yellow are not visible, they are within the model-based simulation (green). A subplot of Figure 6, where a shorter period is zoomed, is added to Supplementary Figure S3.

**Table 4** | Median values of calculated NSEs of the simulated flow (daily and weekly) using the 50 optimum parameter sets for the three forcing datasets (gauge, gridded and model-based) for the three catchments during the calibration (a) and verification (b) period

| | Daily | | | Weekly | | |
|---|---|---|---|---|---|---|
| | Gauge | Gridded | Model-based | Gauge | Gridded | Model-based |
| a) Calibration period | | | | | | |
| Namsvatn | 0.87 | **0.87** | 0.84 | 0.90 | **0.91** | 0.87 |
| Follavatn | **0.65** | 0.64 | 0.60 | **0.76** | 0.73 | 0.71 |
| Tevla | 0.62 | 0.63 | **0.64** | 0.70 | 0.73 | **0.73** |
| b) Verification period | | | | | | |
| Namsvatn | **0.84** | 0.64 | 0.80 | **0.89** | 0.73 | 0.86 |
| Follavatn | **0.69** | 0.64 | 0.56 | **0.79** | 0.71 | 0.67 |
| Tevla | 0.66 | **0.69** | 0.67 | 0.74 | **0.81** | 0.72 |

Maximum NSE among the three forcing datasets is marked in bold.

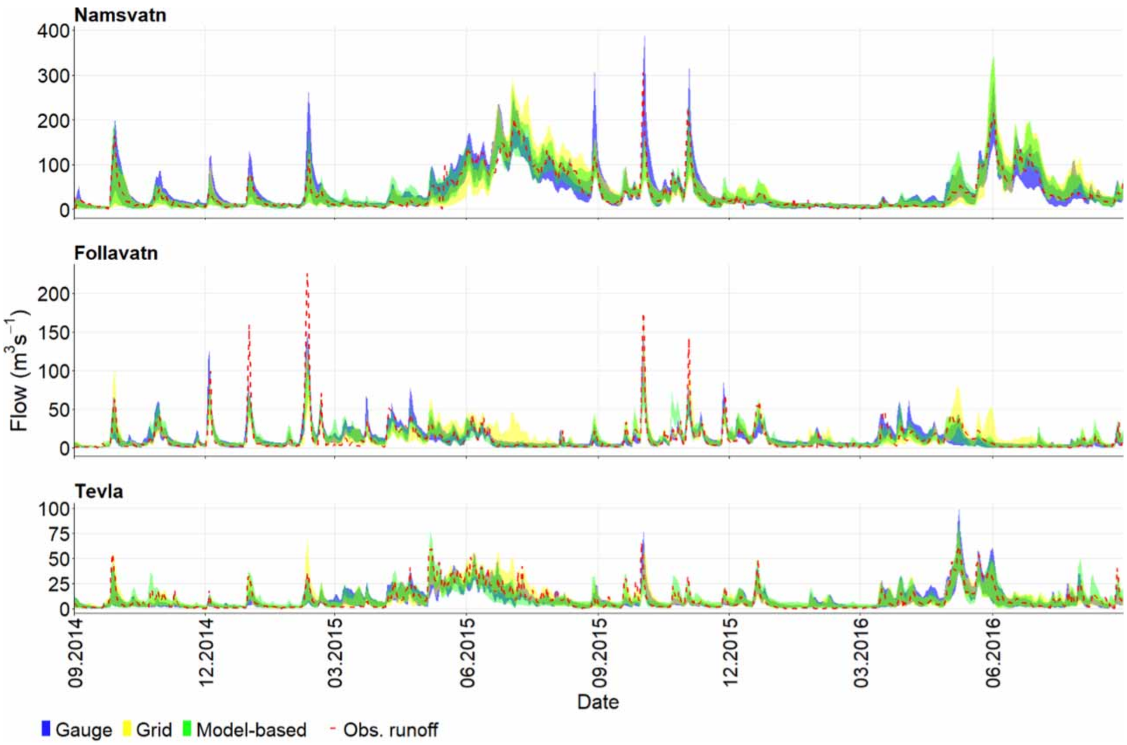


**Figure 6** | Ribbon plot showing three colour bands of maximum and minimum of the 50 simulated flow ensembles for three different forcing datasets for the three catchments during the verification period. Observed runoff is denoted by a red dashed line. Please refer to the online version of this paper to see this figure in colour: http://dx.doi.org/10.2166/nh.2020.076.

Looking at Figure 6, it is visible that the observed runoff mostly falls within the band of maximum and minimum for all three datasets except a few extremes. The bandwidth denotes the uncertainty in the simulated daily river flow due to model parameter uncertainty. A visual inspection of Figure 6 shows that the uncertainty in the simulated flow for the three datasets is generally in the same extent.

Looking at Figure 7, the average simulated annual runoff volume for all three forcing datasets are nearly the same as the observed annual runoff volume, and the difference is less than 15% of the observed runoff volume except for Namsvatn in the year 2012 for model-based data (24%) and Tevla in the year 2014 for gauge data (17%). Further, except Namsvatn in the year 2012 for model-based data, the observed runoff volume falls within the lower and upper value of the simulated flow volume from the 50 parameter sets for all three datasets for all three catchments.

The gridded dataset yielded the best-simulated flow volume compared to observed runoff. The percentage mean absolute error for the annual simulated flow volume with reference to the observed runoff volume is 4–6% for gridded and 5–10% for model-based, while it is 5–9% for gauge dataset.

In Figure 7, the range of 50 simulated flow volumes is shown using the error bar, which represents the uncertainty in the simulated flow volume due to the model parameter uncertainty.

The length of the error bars is between 20 and 30% of observed runoff for Namsvatn and Follavatn, while it is 35–55% for Tevla. Looking at Figure 7, the size of the error bar does not differ largely for the three forcing datasets. The error bars for the gridded dataset are relatively smaller for Namsvatn and Tevla catchments, and they are smaller for model-based data for Follavatn catchment.

## Parameter uncertainty

We investigated the HBV model parameter values of the 50 optimum parameter sets to see how they vary with the forcing dataset. Apart from precipitation correction factor – rainfall (PCORR) and precipitation correction

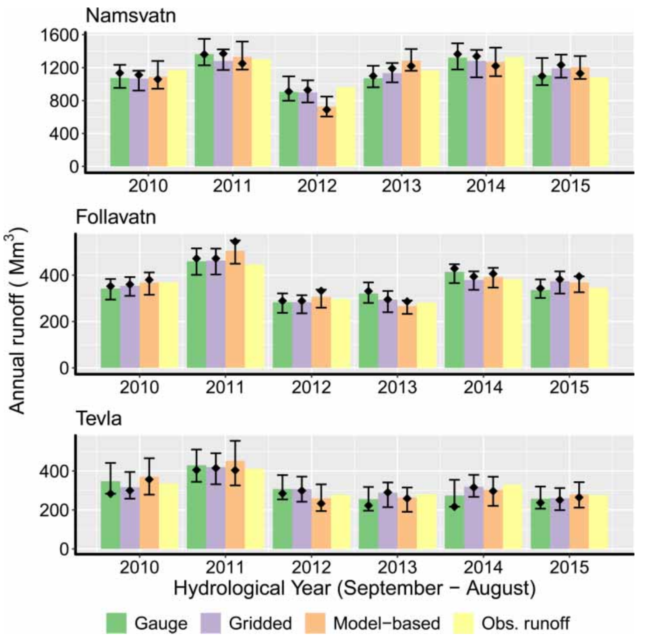


**Figure 7** | Bar plot of the average simulated annual runoff volume of the 50 best MC parameter sets for the three forcing datasets (gauge, gridded and model-based) and observed runoff for the three catchments during the calibration (September 2010–August 2014) and verification period (September 2014–August 2016). The error bar denotes the lower and upper value of the simulated volume from the 50 parameter sets. The simulated volume using the parameter set with the best NSE during the calibration period is denoted by a black diamond point.

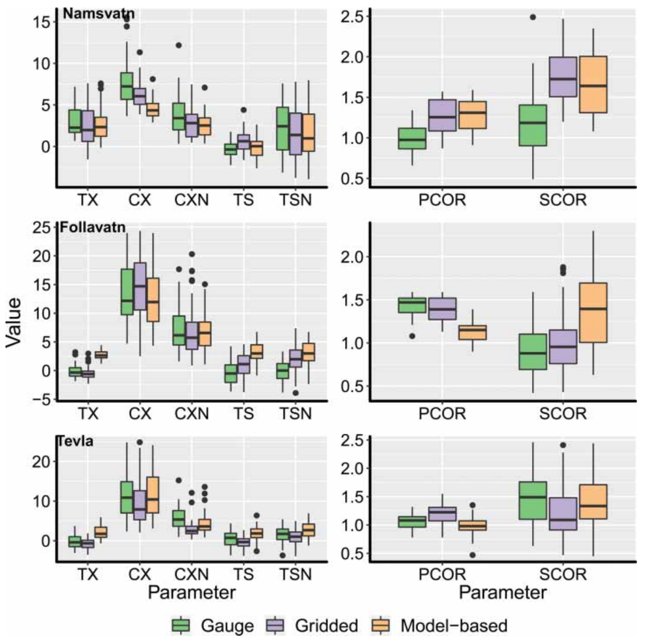


**Figure 8** | Summary of the resulting range of values for the HBV model parameters of the 50 optimum parameter sets for the three forcing datasets for the three catchments. Here, the parameters which showed variation with forcing datasets are only displayed.

factor – snowfall (SCORR), parameters in the snow routine, threshold temperature for rain/snow (TX), degree day factor (CX), degree day factor – forest zones (CXN), threshold temperature for melt/freeze (TS) and threshold temperature for melt/freeze – forest zones (TSN) showed differences in the range of values depending on the forcing dataset for a given catchment, and these parameters are shown in Figure 8. The rest of the calibration parameters assumed a similar range of values for three different forcing datasets.

The calibration parameters, such as PCORR and SCORR, correct the rainfall and snowfall input to the HBV model. This correction for the precipitation input covers several factors, including catch errors and lack of representativeness of gauges (Sælthun 1996). For observational gridded and model-based datasets, a need for the correction can also be due to under/overestimation by the interpolation techniques and data assimilation in NWP models.

In Figure 8, PCORR and SCORR for model-based and gridded data for Namsvatn catchment are similar, and they are higher than the values associated with gauge data. This agrees with the underestimation of model-based and gridded precipitation input compared to gauged precipitation for Namsvatn catchment, as shown in Figure 4. For the Follavatn catchment, PCORR and SCORR for model-based assumed a considerably different range of values to the gauge and gridded dataset, while they were similar to the gauge data for the Tevla catchment, but here different compared to the gridded dataset.

To check whether the estimates of actual evapotranspiration in the model influence the variation of PCORR and SCORR, we plotted the ratio of actual evapotranspiration to precipitation input for the three forcing datasets for the three catchments during the calibration for the best 50 parameter sets (Supplementary material Figure S4). It appears that the ratio of actual evapotranspiration to precipitation is almost the same for all three datasets. It shows that PCORR and SCORR were not influenced by the estimation of actual evapotranspiration in the model, and they vary depending on the forcing dataset.

PCORR and SCORR are of primary concern when a different source of precipitation input is used. The results show that snow routine parameters also highly depend on the forcing dataset. Looking at Figure 8, TX, TS and TSN values for the model-based data were higher than the gauge and gridded for Follavatn and Tevla catchments. It can be seen that CX assumed relatively higher values (10–20) for all three forcing datasets, compared to the traditional range of values for CX (1–5) in the literature (Killingtveit & Sælthun 1995; Lawrence & Haddeland 2011).

Kuczera & Williams (1992) demonstrated that the parameter uncertainty increases with the uncertainty in the areal precipitation input. Looking at Figure 8, the uncertainty (size of the range) in the parameters associated with model-based data is nearly similar in extent to the gauge and gridded datasets.

The calibrated models for each catchment using the three input datasets (50 optimal parameter sets of each atmospheric forcing) were forced with model-based data during the verification period, and the computed NSE is shown in Figure 9. The HBV model calibrated with model-based data performed noticeably better than the model calibrated with gauge or gridded dataset for all three catchments.

With reference to Figure 8, for the Namsvatn catchment, parameters 'PCOR' and 'SCOR' are nearly the same for the gridded and model-based datasets. When the model calibrated with gridded was forced with the model-based dataset, the performance is nearly as good as the model calibrated using the model-based dataset. In contrast, for Follavatn and Tevla, 'PCOR' and 'SCOR' associated with model-based dataset are different from the values obtained using the gauge and gridded datasets and the performance was poorer as shown in Figure 9. This underlines the importance of using the same dataset for the model calibration as is later used in the operational forecasting of inflow.

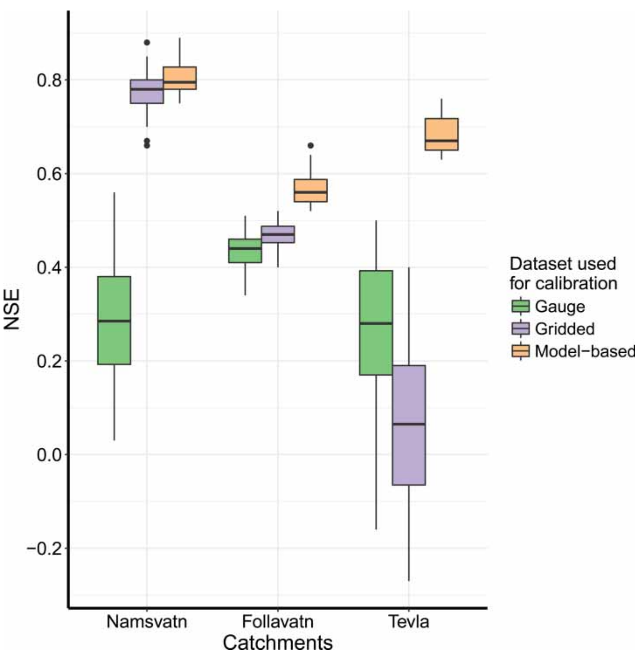


**Figure 9** | Box plot of NSE computed for the simulated flow during the verification period (2014–2016) using model-based dataset as input but using three different sets of the 50 optimal parameters calibrated using the three forcing datasets (gauge, gridded and model-based).

## DISCUSSION

This study investigated the potential of NWP model-based meteorological reanalysis as an alternative to traditional gauge observations for hydrological modelling. In this paper, we showed that the performance of the model-based data was nearly as good or even better than the gauge and observational gridded dataset.

For this assessment, we adopted an MC approach to model calibration (Seibert 1997; Seibert 2003; Steele-Dunne *et al.* 2008; Ledesma & Futter 2017). Even though the MC method with uniform random sampling (100,000 runs in this study) is time and resource consuming, the approach can map most of the feasible parameter combinations; hence, it provides a solid basis for investigating the uncertainty in the model parameter and the response. In this study, we found that the model parameter uncertainty and the uncertainty in the simulated flow using the model-based data as an input was comparable to or even lower than those associated with the gauge and gridded dataset.

Several studies (Te Linde *et al.* 2008; Lauri *et al.* 2014; Essou *et al.* 2016b; Roth & Lemann 2016) assessed the use of reanalysis dataset as an alternative input data for hydrological modelling. This study extends current work with the evaluation of model-based data as an alternative input to hydrological models used in inflow predictions for operational hydropower production planning. In addition, a few studies (Steele-Dunne *et al.* 2008) in the literature focussed on the investigation of model parameter uncertainty depending on different forcing datasets and its consequence in the model response as presented in this paper.

Essou *et al.* (2016b) tested the global and regional reanalysis dataset as an input to a hydrological model in 370 catchments in the United States comparing the output with observational gridded data. Their results showed that the regional reanalysis dataset, which is assimilated using ground-based precipitation observations, produced simulated river flow similar to observed flows. Even though the performance of the global reanalysis was also similar to observed flows, they found that performance was degraded by precipitation seasonality biases.

For well-regulated hydropower schemes, long-term predictions are generally more important than short-term forecasts. Vice versa, poorly regulated schemes are in higher need of short-term inflow forecasts. Such schemes risk flood-spill, when inflow exceeds through flow capacity. Among the three hydropower systems investigated in this study, Namsvatn consists of a large reservoir with high regulation capacity, while the other two catchments have a relatively small regulation capacity. Further, it can be noted that all catchments are snow-fed. Hence, long-term operational planning is required to predict the long-term volume and seasonal distribution based on snow storage in the catchments of all three hydropower systems. In this study, the model-based data simulated the flow volume as well as the observed; hence, it can be a potential alternative to gauge measurements for long-term operational hydropower planning. However, the comparison of model-based data with gauge observations showed that model-based data underestimate the daily extreme precipitation. Moreover, the HBV model that is based on model-based data failed to simulate some of the observed high peaks. This is probably of less importance for the hydropower systems, which consist of large reservoirs with high regulation capacity, than for systems with low regulation capacity.

NWP is an evolving field with the advancement of data science and computer technology (advanced data assimilation techniques). The quality of model-based data products will further improve in the future. Hence, the performance of hydrological model simulations using model-based data products can be foreseen to become better than today in the future.

The calibrated models will be forced by meteorological forecasts (1–10 days) to predict the inflow for short-term hydropower operational planning. We did a verification test on calibrated models using three different forcing datasets and found a clear advantage of using a dataset for calibration which is similar to the source of the dataset being used for the operational simulation. This result implies that the model-based data could be a better alternative for calibrating and updating hydrological models used for inflow forecasting when the forecasting dataset and the model-based dataset come from the same NWP model. Model-based data derived from the same NWP model, which is used to generate meteorological forecasts, will be provided on the same grids and derived using the same methods. Consequently, the data structure of the model-based historical data and the meteorological forecasts would be similar, and the model parameter uncertainty would be similar in the calibration period and in the forecast period. The use of the past records of meteorological forecasts to evaluate predicted flow from models calibrated using the three forcing datasets used in this study is recommended as a future study for this work.

## CONCLUSIONS

The evaluation of model-based input data for hydrological modelling in this study showed that model-based precipitation and air temperature can be a potential alternative to those obtained from gauge measurements and observational gridded data.

The correlation between model-based and gauge data was varying among gauge locations, and the median value of correlation for daily precipitation was 0.8. However, the performance of model-based input data with daily timestep was nearly as good or even better than the gauge or gridded data for the model calibration. It was found that the model parameter uncertainty and simulation uncertainty associated with model-based data appeared as similar to gauge and gridded datasets. Further, the annual simulated flow volume using the model-based data as an input was nearly the same as the observed annual runoff volume.

These results indicate that model-based data can be a potential alternative input to the hydrological models used for inflow predictions for long-term operational hydropower planning. This could be very useful in remote catchments with few gauges and in areas where installing gauges is impossible. Further, results also imply that model-based data can be a promising data source for calibrating hydrological models used for short-term inflow predictions as meteorological forecasts would then have similar sources and similar data structure to the dataset used for the model calibration.

## ACKNOWLEDGEMENTS

The authors thankfully acknowledge the Nord-Trøndelag Elektrisitetsverk (NTE) and Norwegian Meteorological Institute (met.no) for providing hydrometeorological data for this study. The authors thank Martí Garcia Cusí at the Norwegian University of Science and Technology (NTNU) for his initial work on the modelling. The authors gratefully acknowledge the Norwegian Research Council and Norconsult for funding this research work under the Industrial Ph.D. scheme (Project No.: 255852/O30).

## CONFLICT OF INTEREST

The authors declare that there are no competing interests.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this paper is available online at https://dx.doi.org/10.2166/nh.2020.076.

## REFERENCES

Bergström, S. 1976 *Development and Application of a Conceptual Runoff Model for Scandinavian Catchments*, SMHI Rep. RHO 7. Swedish Meteorological and Hydrological Institute, Norrköping, Sweden.

Bergström, S. 1992 *The HBV Model: Its Structure and Applications*. Swedish Meteorological and Hydrological Institute, Norrköping, Sweden.

Berne, A. & Krajewski, W. F. 2013 Radar for hydrology: unfulfilled promise or unrecognized potential? *Advances in Water Resources* **51**, 357–366.

Beven, K. 2012 *Rainfall-Runoff Modelling: The Primer*, 2nd edn. John Wiley & Sons Ltd, Chichester, UK.

Beven, K. & Binley, A. 1992 The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes* **6**, 279–298.

Essou, G. R., Arsenault, R. & Brissette, F. P. 2016a Comparison of climate datasets for lumped hydrological modeling over the continental United States. *Journal of Hydrology* **537**, 334–345.

Essou, G. R., Sabarly, F., Lucas-Picher, P., Brissette, F. & Poulin, A. 2016b Can precipitation and temperature from meteorological reanalyses be used for hydrological modeling? *Journal of Hydrometeorology* **17**, 1929–1950.

Førland, E., Allerup, P., Dahlström, B., Elomaa, E., Jónsson, T., Madsen, H., Perälä, J., Rissanen, P., Vedin, H. & Vejen, F. 1996 *Manual for Operational Correction of Nordic Precipitation Data*. DNMI-Reports 24, p. 66.

Häggmark, L., Ivarsson, K.-I., Gollvik, S. & Olofsson, P.-O. 2000 MESAN, an operational mesoscale analysis system. *Tellus A: Dynamic Meteorology and Oceanography* **52**, 2–20.

Hasan, M. M., Sharma, A., Mariethoz, G., Johnson, F. & Seed, A. 2016 Improving radar rainfall estimation by merging point rainfall measurements within a model combination framework. *Advances in Water Resources* **97**, 205–218.

Haylock, M., Hofstra, N., Tank, A. K., Klok, E., Jones, P. & New, M. 2008 A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research: Atmospheres* **113**, D20119.

Hwang, Y., Clark, M., Rajagopalan, B. & Leavesley, G. 2012 Spatial interpolation schemes of daily precipitation for hydrologic modeling. *Stochastic Environmental Research and Risk Assessment* **26**, 295–320.

Killingtveit, Å. & Sælthun, N. R. 1995 *Hydrology*. Norwegian Institute of Technology, Division of Hydraulic Engineering, Trondheim, Norway.

Kirchner, J. W. 2009 Catchments as simple dynamical systems: catchment characterization, rainfall-runoff modeling, and doing hydrology backward. *Water Resources Research* **45**, W02429.

Kuczera, G. & Williams, B. J. 1992 Effect of rainfall errors on accuracy of design flood estimates. *Water Resources Research* **28**, 1145–1153.

Lauri, H., Räsänen, T. & Kummu, M. 2014 Using reanalysis and remotely sensed temperature and precipitation data for hydrological modeling in monsoon climate: Mekong River case study. *Journal of Hydrometeorology* **15**, 1532–1545.

Lawrence, D. & Haddeland, I. 2011 Uncertainty in hydrological modelling of climate change impacts in four Norwegian catchments. *Hydrology Research* **42**, 457–471.

Ledesma, J. L. & Futter, M. N. 2017 Gridded climate data products are an alternative to instrumental measurements as inputs to rainfall–runoff models. *Hydrological Processes* **31**, 3283–3293.

Lorenz, C. & Kunstmann, H. 2012 The hydrological cycle in three state-of-the-art reanalyses: intercomparison and performance analysis. *Journal of Hydrometeorology* **13**, 1397–1420.

Lussana, C., Ole, E. T. & Francesco, U. 2016 *seNorge v2.0: An Observational Gridded Dataset of Temperature for Norway*. Norwegian Meteorological Institute, Norway.

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 System Science Data* **10**, 235.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I – a discussion of principles. *Journal of Hydrology* **10**, 282–290.

Oke, A., Frost, A. & Beesley, C. 2009 The use of TRMM satellite data as a predictor in the spatial interpolation of daily precipitation over Australia. In: *Proceedings of the 18th World IMACS/MODSIM Congress*, Cairns, Australia, 13–17 July 2009.

Parker, W. S. 2016 Reanalyses and observations: what's the difference? *Bulletin of the American Meteorological Society* **97**, 1565–1572.

Photiadou, C., Weerts, A. & Van Den Hurk, B. 2011 Evaluation of two precipitation data sets for the Rhine River using streamflow simulations. *Hydrology and Earth System Sciences* **15**, 3355–3366.

Rinde, T. 1999 *PINEHBV – A Workbench for Hydrological Modelling*. Nordic Workshop on HBV and Similar Runoff Models. Finnish Environmental Institute, Hydrology and Water Management Division, Helsinki.

Roth, V. & Lemann, T. 2016 Comparing CFSR and conventional weather data for discharge and soil loss modelling with SWAT in small catchments in the Ethiopian Highlands. *Hydrology and Earth System Sciences* **20**, 921–934.

Sælthun, N. R. 1996 *The Nordic HBV Model*. Norwegian Water Resources and Energy Administration, Oslo, Norway, Publication No. 7, pp. 1–26.

Seibert, J. 1997 Estimation of parameter uncertainty in the HBV model: paper presented at the Nordic Hydrological Conference (Akureyri, Iceland – August 1996). *Hydrology Research* **28**, 247–262.

Seibert, J. 2003 Reliability of model predictions outside calibration conditions: paper presented at the Nordic Hydrological Conference (Røros, Norway, 4–7 August 2002). *Hydrology Research* **34**, 477–492.

Sivasubramaniam, K., Sharma, A. & Alfredsen, K. 2018 Estimating radar precipitation in cold climates: the role of air temperature within a non-parametric framework. *Hydrology and Earth System Sciences* **22**, 6533–6546.

Steele-Dunne, S., Lynch, P., Mcgrath, R., Semmler, T., Wang, S., Hanafin, J. & Nolan, P. 2008 The impacts of climate change on hydrology in Ireland. *Journal of Hydrology* **356**, 28–45.

Takahashi, K., Takaya, Y. & Kobayashi, S. 2010 Re-analyses as predictability tools. In: *ECMWF Seminar on Predictability in the European and Atlantic Regions*, Shinfield Park, Reading, 6–9 September 2010.

Talagrand, O. 1997 Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan* **75**, 191–209.

Taskinen, A. & Söderholm, K. 2016 *Operational Correction of Daily Precipitation Measurements in Finland. Boreal. Environ. Res.* **21**, 1–24.

Te Linde, A. H., Aerts, J. C. J. H., Hurkmans, R. T. W. L. & Eberle, M. 2008 Comparing model performance of two rainfall-runoff models in the Rhine basin using different atmospheric forcing data sets. *Hydrology and Earth System Sciences* **12**, 943–957.

Vaze, J., Post, D., Chiew, F., Perraud, J.-M., Teng, J. & Viney, N. 2011 Conceptual rainfall–runoff model performance with different spatial rainfall inputs. *Journal of Hydrometeorology* **12**, 1100–1112.

Villarini, G. & Krajewski, W. F. 2010 Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall. *Surveys in Geophysics* **31**, 107–129.

Vu, M., Raghavan, S. V. & Liong, S. Y. 2012 SWAT use of gridded observations for simulating runoff – a Vietnam river basin study. *Hydrology and Earth System Sciences* **16**, 2801–2811.

Wagener, T., Wheater, H. S. & Gupta, H. V. 2004 *Rainfall-Runoff Modelling in Gauged and Ungauged Catchments*. Imperial College Press, London, UK.

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. *Hydrology and Earth System Sciences* **19**, 951–967.

Yang, Y., Wang, G., Wang, L., Yu, J. & Xu, Z. 2014 Evaluation of gridded precipitation data for driving SWAT model in area upstream of three gorges reservoir. *PLoS ONE* **9**, e112725.

First received 28 May 2019; accepted in revised form 18 December 2019. Available online 22 January 2020