2017:00737- Unrestricted

Report

Application of HYPE in Norway

Assessment of the hydrological model HYPE as a tool to support the implementation of EU Water Framework Directive in Norway

Authors
Lennart Hagen Schönfelder, SINTEF Energi
Tor Haakon Bakken, SINTEF Energi
Knut Alfredsen, NTNU
Abebe Girmay Adera, NTNU

SINTEF Energy Research
Water Resources
2017-12-22

SINTEF Energi AS
SINTEF Energy Research
Address:
Postboks 4761 Torgarden
NO-7465 Trondheim
NORWAY
Switchboard: +47 73597200

energy.research@sintef.no
www.sintef.no/energi
Enterprise /VAT No:
NO 939 350 675 MVA

KEYWORDS:
HYPE
Hydrology
EU WFD
Hydrological indices
Regulated rivers
Hydropower

Report

Application of HYPE in Norway

Assessment of the hydrological model HYPE as a tool to support the implementation of EU Water Framework Directive in Norway

VERSION
Final version

DATE
2017-12-22

AUTHORS
Lennart Schönfelder and Tor Haakon Bakken, SINTEF Energi
Knut Alfredsen and Abebe Girmay Adera, NTNU

CLIENT
Miljødirektoratet

CLIENT'S REF.
Jo Halvard Halleraker

PROJECT NO.
Project No 502001483

NUMBER OF PAGES:
41

ABSTRACT
The precipitation-runoff model HYPE produces relevant hydrological output as support for the implementation of the European Water Framework Directive. The possibility of generating runoff time series for unregulated conditions, automated predictions for ungauged catchments and the implementation of flow alteration modules were identified as model strengths in comparison to other state-of-the-art models. In the present study, HYPE was set up using daily time-steps and was calibrated using discharge time series of 36 gauges for two large domains in southern and central Norway. A split sample validation was executed successfully with an average Kling-Gupta-Efficiency (KGE) of 0.69. Parameter transfer to proxy catchments showed satisfactory results (mean KGE 0.5). Performance weaknesses were identified for peak discharges and low-flow periods. A set of hydrological indices based on simulated natural conditions of water-bodies were calculated for four test cases and compared to those of observed regulated conditions. This can be used to indicate the deviations of the regulated water-bodies' ecological states from its natural conditions.

PREPARED BY
Lennart Hagen Schönfelder

SIGNATURE
L.Sch

CHECKED BY
Hans-Petter Fjeldstad

SIGNATURE
H. Fjeldstad

APPROVED BY
Knut Samdal

SIGNATURE
Knut Samdal

REPORT NO.
2017:00737

ISBN
978-82-14-06594-7

CLASSIFICATION
Unrestricted

CLASSIFICATION THIS PAGE
Unrestricted

MILJØDIREKTORATET REPORT NO.
M-1050|2018

Acknowledgements

This report is the final deliverable from the project titled "Application of HYPE in Norway", being part of the collaboration between Norwegian and Swedish water management authorities, and aimed at supporting the implementation of EU Water Framework Directive in Norway.

The project was carried out under the responsibility of SINTEF Energy Research, with the Norwegian University of Science and Technology (NTNU) and as sub-contractor. The Swedish Meteorological and Hydrological Institute (SMHI) also contributed to the project by providing very valuable scientific support throughout the project, which is greatly appreciated. Our particular thanks go to our WHIST and HYPE modelling experts Kristina Isberg, Joel Dahné and Niclas Hjerdt in Norrköping.

The project was co-funded by the Norwegian Environment Agency (Miljødirektoratet) and the Swedish Agency for Marine and Water Management/SMHI.

Sammendrag på norsk

Denne rapporten oppsummerer erfaringene med anvendelse av den hydrologiske modellen HYPE, utviklet av SMHI i Sverige, med det formål å støtte implementeringen av EUs Vanndirektiv i Norge. HYPE er en prosess-basert, semi-distribuert nedbør-avløpsmodell, og er gratis tilgjengelig for bruk og under kontinuerlig utvikling. Modellen er i Sverige anvendt ved karakterisering av vannforekomster, etablering av miljømål, utarbeidelse av tiltaksplaner og overvåking.

Modellen ble i dette prosjektet satt opp for store deler av fastlands-Norge, fra Nordland i nord til Oslo i sør, fordelt på ett modellområde i sentral-Norge og ett i Sør-Norge. 36 målestasjoner med vannføringsdata med døgnoppløsning ble benyttet til kalibrering og validering, som videre dannet utgangspunkt for regionalisering av parametersett for beregning av vannføring i umålte felt. Modellen simulerte alle nedbørfelt som uregulerte og ga følgelig data for situasjonen før regulering i de felt som idag er regulerte.

Simulerte vannføringsdata dannet grunnlaget for beregning av et sett med hydrologiske indekser, slik som årlig middelvannføring, 7-døgns lavvannføring sommer og vinter. Disse, og en rekke andre, er hydrologiske indekser som alle vurderes å være relevante for å beskrive den økologiske endringen gitt av endringer i vannføring forårsaket av regulering. For å beregne endringer i de hydrologiske indeksene før og etter regulering ble observerte vannføringsdata ble benyttet. Basert på erfaringer i dette studiet konkluderer vi med følgende hovedpunkter:

- HYPE kan simulere vannføring og en rekke andre relevante hydrologiske variable for å støtte implementeringen av EUs vanndirektiv i Norge. Basert på tidsserier av vannføring før og etter regulering kan hydrologiske indekser som anses relevante for å beskrive økologiske endringer beregnes, slik som endring i årlig vannføring, 7-døgns lavvannføring sommer og vinter og endring i flomfrekvens og –størrelse, hvis tilstrekkelig lange tidsserier er tilgjengelig. Beregning av slike hydrologiske indekser krever at modellen er kalibrert med tilfredsstillende prestasjon med hensyn på de situasjoner de hydrologiske indeksene beskriver.

- Resultatene viser at HYPE kan simulere hydrologiske variable med en tilfredsstillende prestasjon, vurdert ved hjelp av det statistiske kriteriet Kling-Gupta-Efficiency (KGE). Kalibreringsresultatene for sentral-Norge og Sør-Norge viste en KGE-prestasjon på henholdsvis 0,73 og 0,72, mens prestasjonene var henholdsvis 0,73 og 0,68 for valideringsperioden. Erfaringene viste at det kan være vanskelig å oppnå gode resultater samtidig for både høye og lave vannføringer, slik at det ved en senere anvendelse kan være fornuftig å introdusere to ulike parametersett for samme modellområde.

- Modellresultatene viste at det er mulig å tilpasse HYPE til et stort område med en tilfredsstillende prestasjon ved bruk av vannføringsdata fra flere målestasjoner. Dette muliggjør simulering av vannføringer i umålte felt med akseptabel usikkerhet.

- HYPE har innebygde beregningsrutiner som muliggjør en representasjon av vannkraftproduksjon og magasindisposisjoner med times-oppløsning. Dette innebærer at modellen kan simulere vannføring også i regulerte vassdrag. HYPE kan også modellere vanntemperatur og vannkjemi.

- HYPE tilbyr en fleksibel romlig oppløsning og inndeling slik at den er egnet til å beregne vannføringsdata og hydrologiske indekser tilpasset vanndirektivets behov.

Ved en framtidig anvendelse av HYPE som støtte for gjennomføringen av EUs vanndirektiv i Norge anbefaler vi følgende forbedringer i oppsettet av HYPE:

Romlig inndeling/representasjon:

Manuell korrigering og inndeling i delfelt, innlegging av innsjøer og korreksjon av strømningsretning ("flow routing") er en tidkrevende operasjon og bør gjøres samtidig for hele Norge. Inndelingen av delfelt og elver/innsjøer bør gjøres i samsvar med den romlige inndelingen av vannforekomster slik at modellresultatene kan kobles direkte til vannforekomster.

Kalibrering med hensyn på bruk av resultatene:

Studiet viste at det er vanskelig å få gode kalibreringsresultater for både høye og lave vannføringer med samme parameteroppsett. Vi anbefaler derfor at formålet med modelloppsettet avklares tydelig før kalibrering (hvilke hydrologiske indekser som skal beregnes), og at det eventuelt utvikles to ulike parametersett for henholdsvis høye og lave vannføringer.

Landsdekkende modelloppsett:

En av styrkene til HYPE er at den støtter kalibrering basert på flere målestasjoner og automatisk overføring av parametersett. Det er derfor hensiktsmessig å tilpasse modellen samtidig for store områder. Parametertilpasning for nedbørfelt med ulike karakteristika sikrer derfor en robust overføring av modellparametere til umålte felt.

Forbedret datagrunnlag:

Vi anbefaler at det hentes inn data fra Corine Land Cover 2012 -datasettet (http://land.copernicus.eu/pan-european/corine-land-cover/) og Harmonised World Soil Database (http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/), og eventuelt kobler disse datakildene. Dette vil trolig redusere parameterusikkerhet, bedre forholdet mellom prestasjon og modellkompleksitet og redusere arbeidet med tilrettelegging av data og modellkalibrering. Bruk av flere klimavariable som for eksempel stråling vil trolig også forbedre tilpasning av modellen.

Bruk av HYPE til å beregne hydrologiske indekser:

Erfaringene i dette studiet tilsier at HYPE krever så høy kompetanse at det er hensiktsmessig at modellen opereres kun av noen få, dedikerte eksperter. Vi anbefaler derfor at et eventuelt videre arbeid med HYPE sentraliseres til en eller noen få personer og kun modellresultater, i form av ferdige, beregnede hydrologiske indekser, og gjøres tilgjengelig for regional forvaltning gjennom Vann-Nett.

SINTEF

Table of contents

1 Introduction

1.1 European Water Framework Directive (EU WFD) and hydrological alterations

The EU Water Framework Directive (EU WFD) was adopted in 2000 and commits European Union member states to achieve good ecological status or good ecological potential in all water bodies (Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy 2000). EU WFD was transposed into the Norwegian Regulation on a Framework for Water Management, normally referred to as Vannforskriften (The Water Regulation), entered into force in 2007. Norway has taken full part in the Common Implementation Strategy (CIS) for the EU WFD since it was established in 2001 (Common Implementation Strategy for the Water Framework Directive (2000/60/EC)).

EU WFD specifies biological elements for the classification of ecological status of rivers and lakes. These biological quality elements are supported by a set of hydro-morphological, chemical and physio-chemical quality elements. The hydrological regime is one of the hydromorphological quality elements for rivers defined in the EU CIS Guidance document No 13 for the Common Implementation Strategy for the Water Framework Directive (2000/60/EC).

The structure and persistence of biotic communities within river ecosystems are strongly influenced by both spatial and temporal variation in environmental conditions (Poff, Ward 1989; Stanford et al. 1996), where the flow regime is a key component. The most common cause of changes in flow regimes is the storage and regulation of water by dams and reservoirs for the supply of water for human use (Nilsson et al. 2005). Reservoirs can be managed to meet a range of human needs, such as irrigation, domestic water supply, flood mitigation and hydropower generation. In Norway, most of the large river regulations are made for the primary purpose of hydropower production.

The changes in flow introduced by river regulations will vary between the sections of the water course affected by the regulation. A typical hydrological fingerprint of a regulation will be that the magnitude and frequency of floods will be reduced in all downstream areas of the reservoir. As a result of a higher consumption and energy demand, the winter flow downstream of hydropower plant outlets will typically be higher (and warmer) than before regulation. In bypass sections, the annual average flow will be reduced, but the changes introduced are also highly site-specific given by the actual regulation and climatic region.

In assessing hydrological change, Richter et al. (1996) proposed a set of 64 indicators of hydrologic alteration statistics that could be used to quantify the degree of flow regime change on the basis of ecologically meaningful metrics, and the more updated review of Poff and Zimmermann (2010) provides interesting reading related to the topic. Common to the method of hydrologic alteration statistics is the focus on distinctive components of the flow regime, defined in terms of magnitude, timing, duration, frequency, and rate of change. By assessing the changes in the hydrological regime, e.g. expressed by a set of hydrological indices, the impacts of the regulation on various biophysical processes can be assessed, and the changes in hydrological indices can also form the basis for which components that should be restored as part of a river basin management plan. There is on-going research in Norway to define a hydromorphological classification system in line with the principle of the EU WFD (Harby et al. 2017, draft), which will also include a set of hydrological indices.

1.2 Goal of the study

In Sweden, the software programme HYPE is used to provide data to water authorities for the implementation of both the EU WFD and the EU Marine Strategy Framework Directive. The model supports authorities in characterization of the status of water bodies, establishment of environmental goals, remedial measure planning and in development of monitoring strategies. HYPE is also used as a flood forecasting system.

Since Sweden's monitoring network only consists of 300 gauging stations covering approximately 450 000 km^2, there are no monitoring system in place for the vast majority of the water-bodies. The Swedish HYPE model is partly replacing the monitoring system with modelled daily time series for runoff and monthly predictions of nutrient loads. Daily updated HYPE model results are publicly available in an end-user friendly way. Water-body delineation, model results and performance criteria can be viewed on https://vattenwebb.smhi.se/.

Based on the similarities with sparse hydrological monitoring, the need for hydrological data to support the implementation of the EU WFD and the positive experiences with HYPE in Sweden, a project with the following objectives was carried out:

- Evaluate HYPE as a tool for the implementation of the WFD in Norway in terms of effectiveness, applicability and performance
- Investigate if HYPE can be set-up efficiently for a large domain containing several catchments in Norway
- Identify modelling barriers, such as lack of data availability or software tools, data compatibility and insufficient description of hydrological processes in Norwegian conditions
- Demonstrate the generation of hydrological indices for selected water-bodies and evaluate if HYPE is suitable to create information about hydrological reference conditions

2 Methodology

2.1 Description of HYPE

HYPE is a process-based semi-distributed rainfall-runoff model which has been developed at SMHI (Swedish Meteorological and Hydrological Institute) from 2005 to the present. Its code is written in FORTRAN and the software is open source under the Lesser GNU Public license (Free Software Foundation 2007). The open source availability was chosen to initiate and strengthen international collaboration in hydrological modelling.

HYPE is based on HBV (Lindström et al. 1997) and has its main advantages in prediction of discharge in ungauged basins and water quality modelling. The model is under continuous development and this report might possibly not contain information about the very most recent modules and features. An up-to-date comprehensive description of the features, process modules and model structure can be found on the HYPE wiki (http://www.smhi.net/hype/wiki/doku.php). For a complete methodology of the model set-up for central Norway see Schönfelder (2017). The model set-up for central and south Norway in this report used the simplest modules to describe the hydrological processes that are available in HYPE.

Model concept

The catchments are divided into sub-basins that are linked in a horizontal flow network. In turn, the sub-basins are divided into classes, which are not coupled geographically within the sub-basin. The classes consist of a land use and a soil type (SLC – *"soil and land use class"*). Model parameters can be associated with land use, soil type, SLC or be general for the whole catchment or domain respectively. Many process modules in HYPE are similar to the processes in HBV.

The properties of the soil and land use classes define the hydrological reaction of the land surface. Most of the parameters are dependent on either land use or soil type. An SLC can consist of up to three soil layers, the processes include macropore flow, groundwater flow and surface runoff. Example parameters for the soils are field capacity, porosity and wilting point. Processes such as evaporation and snow melt are based on air temperature (day degree method) and are parametrized based on the land use.

Lakes and reservoirs above a selected size threshold are linked within the flow network as a sub-basin with special properties. The lakes' outflow is calculated by either an individual or general rating curve. Lake volume and mean depth can be given as input data. Lakes and reservoirs below the chosen threshold size are calculated as a special SLC as a fraction of a sub-basin and hence cannot have regulation routines. As an example, Figure 1 shows the sub-basin structure of the Gaula catchment in WHIST (SMHI 2016) where the orange sub-basin is selected, upstream basins are depicted in blue, downstream basins in green and tributaries in yellow.

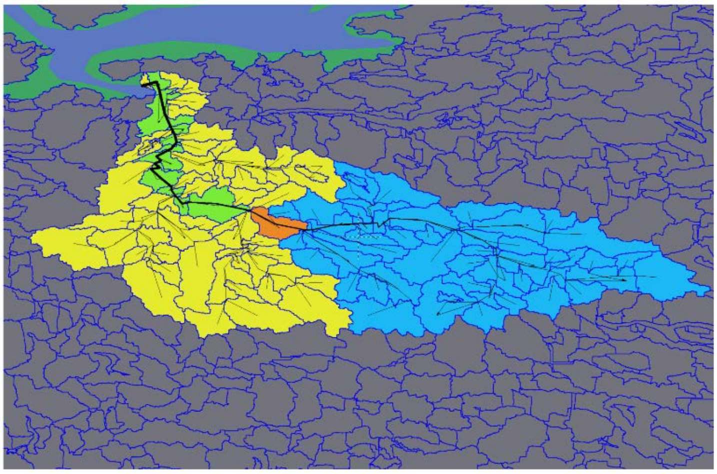

Figure 1 Flow network of the river Gaula in WHIST.

Output

Runoff can be calculated for the outlet of the sub-basins and can therefore be linked to the waterbody that is delimited by the according sub-basin in HYPE. Evaporation, potential evaporation, snow water equivalent and groundwater level are averaged for the sub-basins. Model output is generated as a text file of time series of sub-basin averaged values. Time series for these parameters can be generated for all sub-basins. Additionally, a simulation assessment for all sub-basins where observed data is available can be generated. This assessment yields performance criteria metrics.

Input

The model setup with the simplest modules employed requires daily means of precipitation and temperature of all sub-basins as input data. Since precipitation and temperature time series were obtained as NetCDF (Network Common Data Form) files, they had to be averaged for each sub-basin and timestep with an R-script. HYPE can be run in any timestep, hourly or daily time steps were tested in previous model setups.

Input time series are stored column-wise (each column represents one sub-basin) in a text file. Temperature and precipitation values can therefore be changed e.g. according to season, latitude, longitude and elevation, facilitating the adaption of input data to climate change scenarios.

Model set-up tools

Any GIS tool can be used for the preparation of spatial data (e.g. ArcGIS, QGIS, GrassGIS). The division of the landscape into a linked network of sub-basins can be done using WHIST and a flow accumulation raster map. WHIST also calculates land use and soil classes, elevation and slope for individual sub-basins based on chosen raster data sets. For further processing of geographic data, visualization of the model set-up as well as analysing model results, the comprehensive R package "HYPE tools" is available (http://hypecode.smhi.se/open-source/tools-2/).

2.2 Additional Model features

The following model features were not tested within the scope of this project, but are relevant to support the implementation of the EU WFD and can potentially be used in the future model setups:

Regulation of reservoirs and water transfer

HYPE has several modules to simulate water management. The two most relevant modules for the simulation of regulated reservoirs and water diversion are explained in this sub-section. These modules were not evaluated within this project, but their functionality is demonstrated in the model setup for Sweden and was verified through personal communication (Niclas Hjerdt, SMHI).

The Bifurcation function (1) enables water transfer to a downstream sub-catchment based on discharge time series, a fraction of the outflow of the source sub-catchment, a maximum or a minimum flow. Both donating and receiving sub-catchment can be of any type.

The Management function of HYPE (2) can use discharge time series to transfer water from a reservoir to any other sub-catchment. It is defined as a demand-sided transfer, but water is only transferred if it is available in the source sub-catchment. Only one transfer is possible when time series are used, the function includes a delay of one timestep for the water to arrive in the destination sub-basin.

We assume that a combination of both functions can be employed to simulate both water abstraction from a reservoir for hydropower production and the residual flow in the downstream reach of the reservoir.

The potential implementation of the Bifurcation function (1) for the residual flow reach and the Management function for the hydropower bypass are shown exemplary for the case of the reservoir Innerdalsvatnet and its respective hydropower plant *Brattset* in Figure 2. The sub-basins are shown with black outlines.

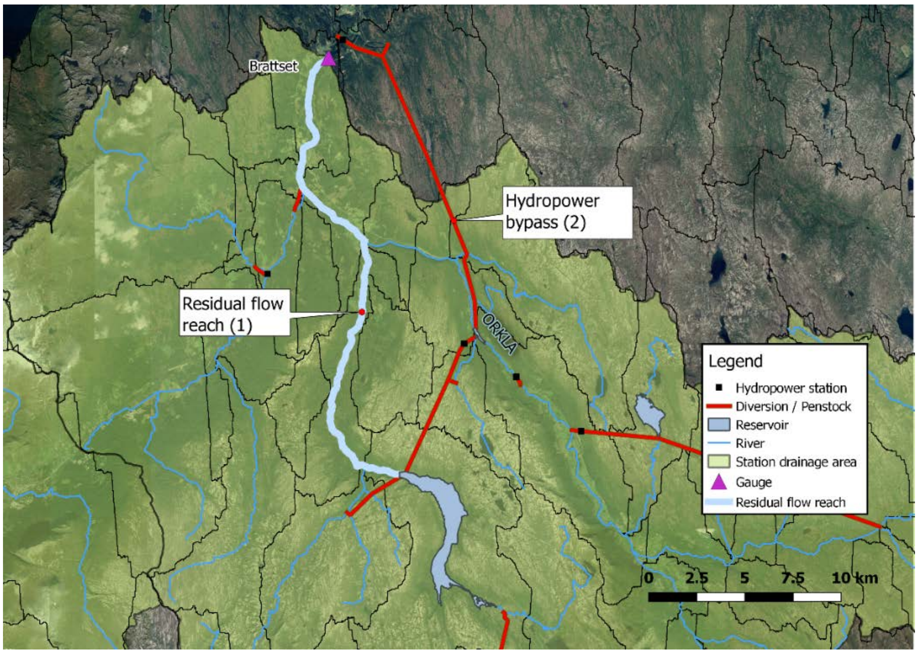

Figure 2 Potential application of two hydropower modules shown for an example case in Orkla catchment

Water temperature and ice conditions

The current input data in the created model setup within this project is sufficient to include a water temperature and surface water ice model, but measured water temperature time series are needed for calibration. Water temperature is calculated as a tracer, and different sources and two calculation modules can be implemented. The model enables heat transfer to the solid soil matrix and atmosphere and includes various heat sources.

Water quality

HYPE simulates nitrogen and phosphorus in soil and water based on processes that include agricultural sources, point sources (e.g. water treatment plants) and atmospheric deposition.

Sediments

The sediment model calculates suspended sediments from soil erosion and algae production in lakes and rivers. Two alternative soil erosion models can be utilised.

Modules

It is possible to choose from various modules that describe the hydrological processes in HYPE. Employing additional input data such as radiation, daily temperature maximums and minimums and snow depths, more comprehensive formulas can be chosen that may lead to better model performance.

2.3 Indices for hydrological alterations

The HYPE project primarily focusses on the sub-element *Quantity and dynamics of water flow* as a hydromorphological quality element of EU WFD. Indices for hydrological alterations (IHA) make it possible to quantify the reference flow conditions and the deviation of altered conditions. Also, they are implemented as proxy metrics to indicate the ecological status of regulated rivers. Furthermore, the difference of hydrological indices between reference conditions and actual flow regime can help to identify the most suitable counter-measures against ecosystem degradation.

The selected statistical parameters in this project are calculated from daily discharge time series. They are based on previous studies that employed IHA (Forseth, Harby 2014; Richter et al. 1996; Hohl 2003). Annual median flow and lowest weekly average flow in summer and winter furthermore link hydromorphological indices to ecosystem indicators such as salmonids (Forseth, Harby 2014).

Annual median flow

Water availability is the basis for any aquatic ecosystem. Discharge is directly related to flow velocity, water-covered area and water depth (Acreman et al. 2009). In bypass sections, where regulation due to hydropower results in a small residual flow, habitats may change their characteristics completely. Similar effects can occur when water is transferred between river catchments. The corresponding hydrological changes can be partly assessed with the annual mean flow.

Low-flows

Extended artificial low-flow periods are limiting aquatic life and may negatively affect connectivity, fish migration and survival. Different statistical parameters for low-flows have been widely accepted as indicator for alteration (Forseth, Harby 2014; Richter et al. 1996).

- Lowest weekly average flow (summer)

The average runoff of the 7-day-period with the lowest accumulated runoff from November to March.

- Lowest weekly average flow (winter)

The average runoff of the 7-day-period with the lowest accumulated runoff from April to September.

- Common low flow (according to Norwegian definition) ("alminnelig lavvannføring")

For the calculation of the common low flow index according to the Norwegian definition, daily runoff is sorted from the highest to the smallest value. Runoff no. 350 is sorted out and with the values from all the years a new series is built. From this series, the average of the higher 2/3 is built (Tallaksen, van Lanen 2004).

Number of rises

Ecological status classification based on different flood frequencies and magnitudes is an expert proposal for salmon rivers (Forseth, Harby 2014). The classification system proposed in Harby et al. (2017) is based on qualitative expert judgment. In this work, we quantify flood frequencies based on the number of periods with uninterrupted increase in runoff over 50 % of the average annual runoff per day (Hohl 2003). This natural number of periods per year is then averaged over the total period.

More IHAs are calculated in sub-section 4.5, for their description we refer to Richter et al. (1996).

2.4 Comparison of hydrological models

In order to set HYPE in a broader model context, we compared the model with a selection of other state-of-the-art precipitation-runoff-models (Table 1).

The compared models are all physically based and conceptual to an extent. Also, they are flexible in terms of spatial and temporal resolution. All models have been set up for large domains in various countries on different continents.

Table 1 Comparison of state-of-the-art hydrological models

	HYPE	Enki	HBV	WEAP
Emphasis / Purpose	- Representation of hydrological processes - Water quality - Predictions in ungauged basins	- Representation of hydrological processes - Research tool	-Representation of hydrological processes	- Water use management - Scenarios for Climate change and water use
Further development	SMHI Team	Small group of individuals	Large user community	Large user community
Hydropower module	- Rating curve for reservoirs - Dynamic Inter-basin transfer - Reservoir runoff time series	-	-	- Rating curve - Constant environmental outflow of reservoirs - Hydropower output
Water temperature module	x	x	-	x
Water quality module	x	-	-	x
Sediments module	x	-	-	-
Automatic calibration	x	x	x	x
Multigauge calibration	x	x	x	x
Parameter transfer	Land use and soil based - automatic	Regression based - manual	Regression based - manual	Land use and soil based - automatic

3 Data and modelled regions

3.1 Regions modelled

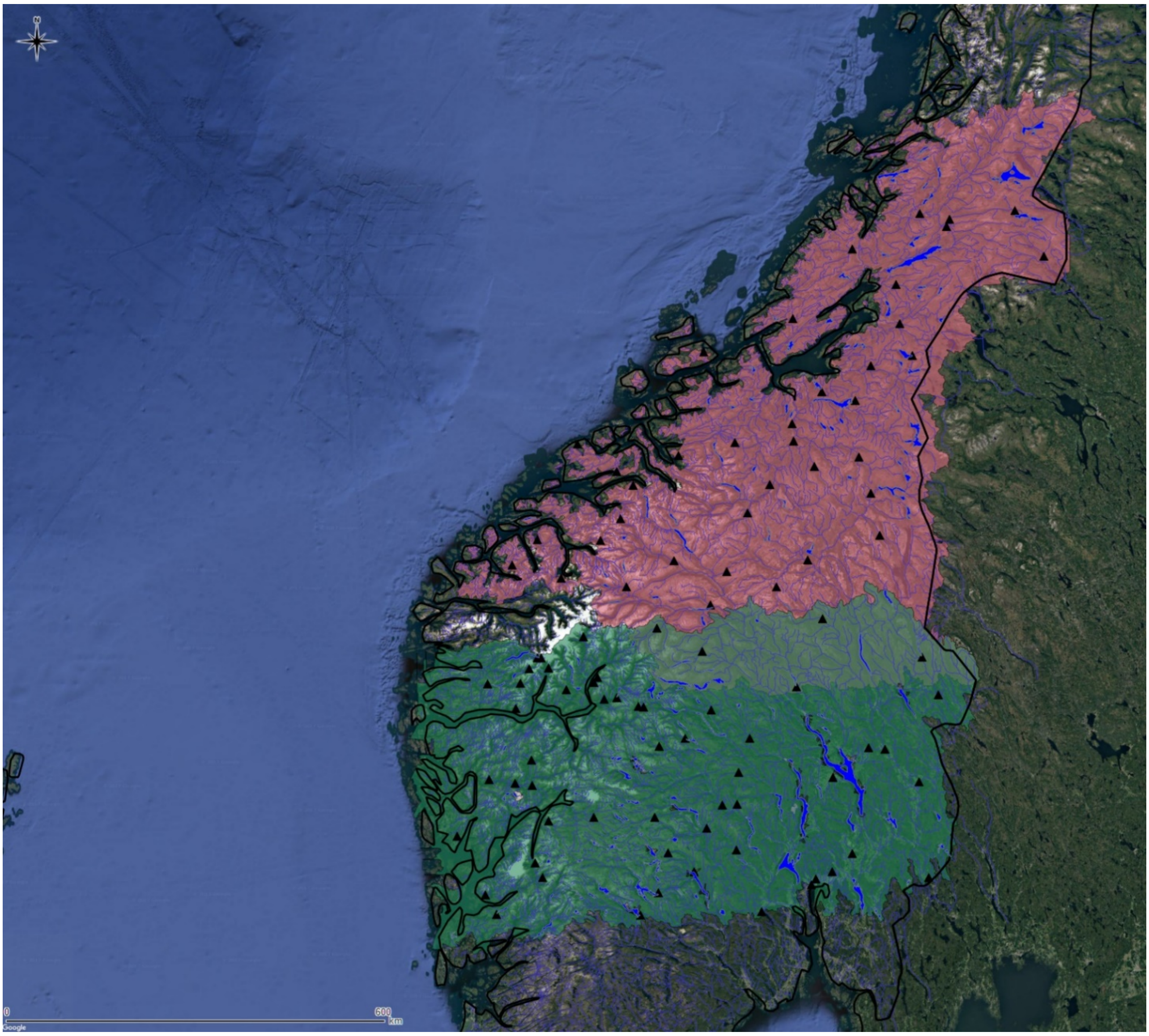

Figure 3 The two modelled regions shown in red and green, with implemented lakes in blue colour and gauging stations depicted as black triangles.

The regions modelled are shown in Figure 3. In the northern region, the testing was focussed on the parameter transfer from five calibrated catchments to five uncalibrated proxy catchments to evaluate the predictive power of HYPE for ungauged rivers.

In the southern region, we focussed on using 31 gauges for calibration to evaluate if a multigauge-calibration can give accurate results consistently for catchments of different sizes, land uses, soils, topology and other characteristics.

3.2 Delineation of catchments

The delineation of sub-basins and their routing amounted for a large share of the project work. WHIST is designed to work with international ready-made databases such as USGS Hydrosheds (https://hydrosheds.cr.usgs.gov/). The most relevant Hydrosheds products are the preconditioned DEMs and their resulting Flow direction and Flow accumulation rasters. They are not available further North than 55° Latitude and therefore do not cover Norway.

WHIST delineates the sub-basin based on the Flow accumulation raster, which was from an originally unconditioned digital elevation model (DEM) of Norway with a resolution of 50 m. The delineation was therefore inaccurate for many catchments, which can have large negative impacts on model results, especially for small catchments. Forced points can be defined before the delineation process to predefine outlets of upstream sub-basin (this can be done with any other location of special interest, e.g. confluence points, hydropower-plants). This was done for all gauges used for calibration and parametrization. As mentioned in sub-section 2.1, lakes above a threshold size are considered as individual sub-basin. They are inserted as a shapefile and "cut" into existing basins. The insertion results in small basins next to the lake and unconnected basins. These were merged and respectively reconnected after the insertion.

Results extraction according to sub-basin system

The most relevant hydrological information for the river water-bodies considered in the EU WFD is river run-off at the up- and downstream boundaries. The delineation of the current model set-up done with WHIST is restricted by size only, the sub-basins are therefore not congruent with the water-bodies in the WFD in most cases. Discharge is calculated at the outlet of HYPE's sub-basins. For the extraction of model results at any given location within the model domain, the closest sub-basin outlet must be chosen.

Figure 4 shows an example of a catchment within the test domain. The green water-bodies show the REGINE system, it has a finer spatial resolution than the sub-basin system in the model setup. The respective downstream boundaries of REGINE are depicted as black triangles. The river (shown as blue line) separates the water-bodies in REGINE enhet, which is a main reason against using it as a basis for the sub-basin delineation for HYPE. The model structure does not allow for parallel catchment areas that are separated by the river.

If extraction of results is done using the current set-up, it must be ensured that the relative difference in catchment area between the basin outlet in HYPE and the location of interest is not exceeding certain thresholds (e.g. 10 %).

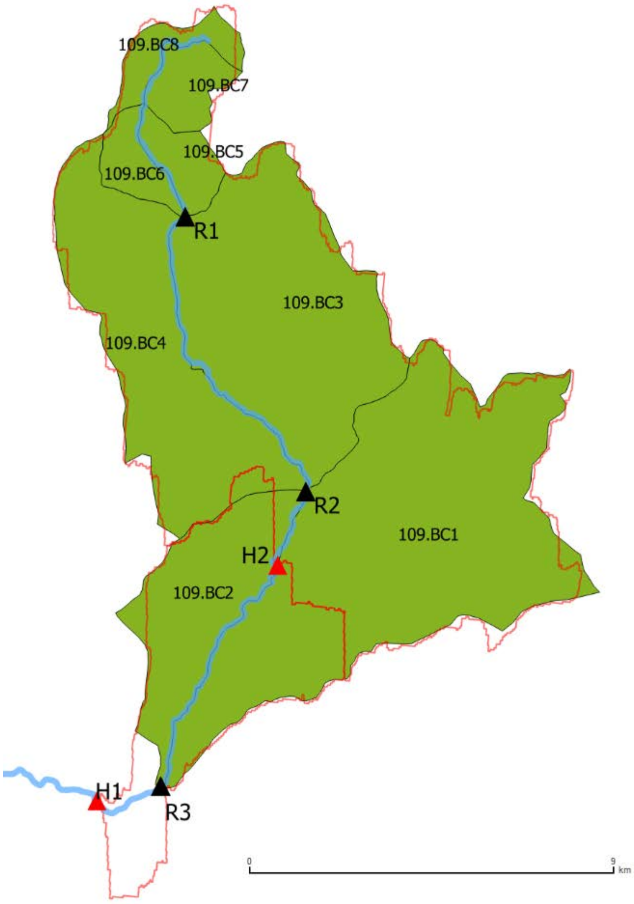

Figure 4 Sub-basin delineation of HYPE (red) and water-bodies in Regine enhet (green with unit IDs).

If the outlet location R1 of water-bodies 109.BC5 and 109.BC6 was relevant, HYPE could not deliver reasonable results, since the most fitting sub-basin outlet point H2 drains a much larger area. This can be expected, since the linear distance R1-H2 is large. If discharge information at R2 is of relevance, H2 is also the most fitting. The error would be less than in the prior case, but H2's drainage area is larger than R2's. In this case, the large drainage area difference is less expected, because the linear distance between H2 an R2 is small. For R3, the linear distance and the drainage area difference to H1 are negligible, therefore the HYPE model results can be applied without any problems. With the current model set-up, it is therefore not guaranteed to get good model results for any given location of interest. The potential challenge is less problematic in larger catchments.

The average sub-basin size in the central Norway setup is 35 km², the average size for south Norway is similar. Results can therefore be extracted every $\sqrt{35\ \mathrm{km}^2} \approx 6\ km$ on average.

3.3 Data and material

Table 2 shows all data sources relevant for the delineation of the sub-basins in WHIST.

Table 2 Databases and files used for delineation of sub-basins.

Data	Name / Type	Data type	Source
Digital elevation model	Digital terrengmodell 50 m	Rastermap	http://data.kartverket.no/download/
Gauge location	Måleserier	Shapefile (points)	https://nedlasting.nve.no/gis/
River network	Elvenett	Shapefile (polyline)	https://nedlasting.nve.no/gis/
River catchments	Nedbørfelt til hav	Shapefile (polygon)	https://nedlasting.nve.no/gis/
Gauge drainage area	Totalnedbørfelt til målestasjon	Shapefile (polygon)	https://nedlasting.nve.no/gis/

Table 3 shows the data employed for creating the SLCs. Entries in *italic* were employed in the current model setup. The given alternatives (Corine Land Cover and Harmonised World Soil Database) are recommendations for future model set-ups. The reasoning behind the recommendations is described in sub-section 4.4.

Table 3 Data used for Soil and Land use classes.

Data	Name / Type	Data type	Source
Land use	Nevina	Rastermap	pers. Communication w/ Astrid Voksø (NVE)
	Corine Land Cover	Rastermap	http://land.copernicus.eu/pan-european/corine-land-cover/
Soil type	Løsmasse	Rastermap	*http://www.ngu.no/*
	Harmonised World Soil Database	Rastermap	http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/

The input and calibration data and their respective sources are shown in Table 4.

Table 4 Input and calibration data.

Data	Name / Type	Data type	Source
Precipitation	Daily sum	NetCDF	http://www.senorge.no/
Temperature	Daily average	NetCDF	http://www.senorge.no/
Observed discharge	Daily average	Comma separated values (csv)	HYDRA II http://www4.nve.no/xhydra/

4 Model results

In sub-sections 4.2 and 4.3, the calibration process and the results are described. Multi-gauge calibrations were executed in both domains, meaning that parameters that describe the physical processes for all catchments were simultaneously calibrated. It is a commonly used approach in HYPE to cover several catchments within one model setup. The advantage over individual calibration is a better model economy, less calibration effort and the generation of a parameter set that is applicable for a large region, also where no calibration was possible because of missing observed discharge data. The derived parameter set is suitable for predictions in ungauged catchments. The disadvantage of this approach is a weaker individual station performance of the model in comparison to individual catchment calibration. Individual catchment calibrations are also possible in HYPE, given that observed discharge data is available.

In sub-section 4.4, the parameters are transferred to uncalibrated catchments.

The model results are derived from a test set-up for HYPE. Performance weaknesses described in the following sub-sections are not necessarily model based. We expect that improved calibration can be achieved using better datasets and by employing the gained experience through this calibration. For many of the calibration weaknesses, the physical background problems and assumed potential resolutions are known.

4.1 Calibration and validation

The calibration goal was to achieve an overall good performance of the model. The Kling-Gupta-Efficiency (KGE) was chosen as the objective function for both regions. The model performance metric KGE was developed as an alternative to NSE and is calculated from observed and simulated discharge time series as follows:

$$KGE = \sqrt{(r-1)^2+(\alpha-1)^2+(\beta-1)^2}$$

r corresponds to the linear correlation coefficient between the observed and simulated flows, α to the ratio of the standard deviations between the observed and simulated flows and β to the bias error between the observed and simulated flows. KGE is developed from the NSE and can reach values from $-\infty$ to 1.

For KGE = 1, simulated values fit observed values perfectly. If the KGE is zero, the model predictions are equivalent in goodness-of-fit to the mean observed value. For KGE < 0, the model results can generally be seen as not satisfactory.

Patil, Stieglitz 2015) implied that KGE values > 0.6 can be seen as satisfactory. For more details on the KGE, we refer to (Zhu, Chen 2016; Gupta et al. 2009).

HYPE was calibrated both automatically and manually. First, manual trial-and-error calibrations were used to identify important parameters. The following automatic calibration routines were employed thereafter:

- The Monte Carlo method uses a large number of model runs with random sampling of the parameters within a parameter range. The number of model runs for this method was set to 100. It was used for testing single parameters and when sensitivity and a meaningful parameter range were not known. The parameter range was set to a high range for the latter case.

- The Differential Evolution Monte-Carlo method DE-MC combines the concept of Markov Chain Monte Carlo and Differential Evolution (DE). It runs several Markov chains in parallel and uses its previous steps to generate the next candidate. The DE defines the scale and orientation of the jumping distribution of the Markov-Chain. It uses the difference of two randomly chosen parameters of the parent generation to find a step size for the next mutation. The number of generations was set to 40 and the number of parallel model runs to 20. This resulted in a calibration run time of 1 hour.

In the following sections 4.2 and 4.3, model results are explained with respect to model parameters and static databases such as soil and land use distribution. Uncertainty analysis of the results and of input data for precipitation and temperature are not discussed in detail within the scope of this project.

4.2 South Norway

For south Norway, three different model setups with increasing numbers of gauges (5, 16 and 31) for calibration were created. In the following discussion, the version with 31 gauges is discussed, since it is the latest version and most suitable for the nationwide approach for Norway. The gauges were calibrated for a seven-year calibration period with the median KGE of all stations as objective function. Initial values for calibration were transferred from HBV (Hveding et al. 1992).

The calibration approach was to calibrate groups of parameters sequentially. Each group corresponded to a physical process, in order to avoid that the calibration of one process compensates for the mechanisms of another process. The groups were sorted and calibrated in the following order:

1. Precipitation
2. Snow melt and snow accumulation
3. Evapotranspiration
4. Infiltration and soil storage
5. Lakes and river routing
6. Soil freezing

The model calibration for southern Norway underestimates peak discharges to some extent.

The overall performance for this calibrated region can be described as satisfactory in terms of the Kling-Gupta efficiency. The average values of KGE 0.72 and KGE 0.68 for the seven-year calibration and validation period of all 31 gauges show the validity of the model.

The multi-gauge calibration of 31 stations and the respective results show the strength of the model in terms of large-domain modelling and transferability of parameters. The risk of over-parameterisation to a specific type of catchment response characteristics is low.

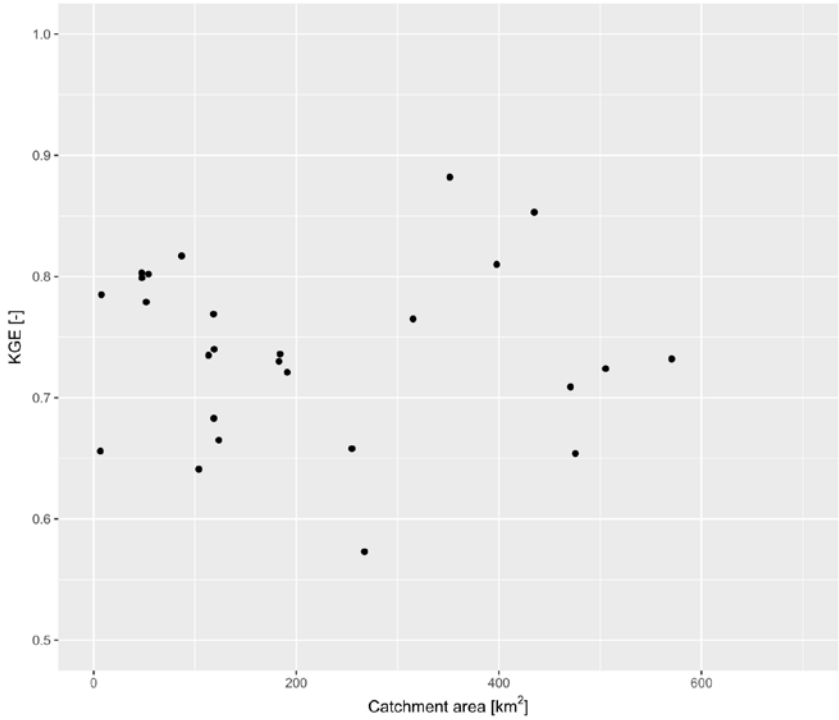

Figure 5 Comparison of KGE and catchment size of 27 gauges.

The preference of good performance for large catchments as shown for central Norway is not prevalent for this model calibration. The scatter plot in Figure 5 shows an equal distribution of KGE over the catchment size.

4.3 Central Norway

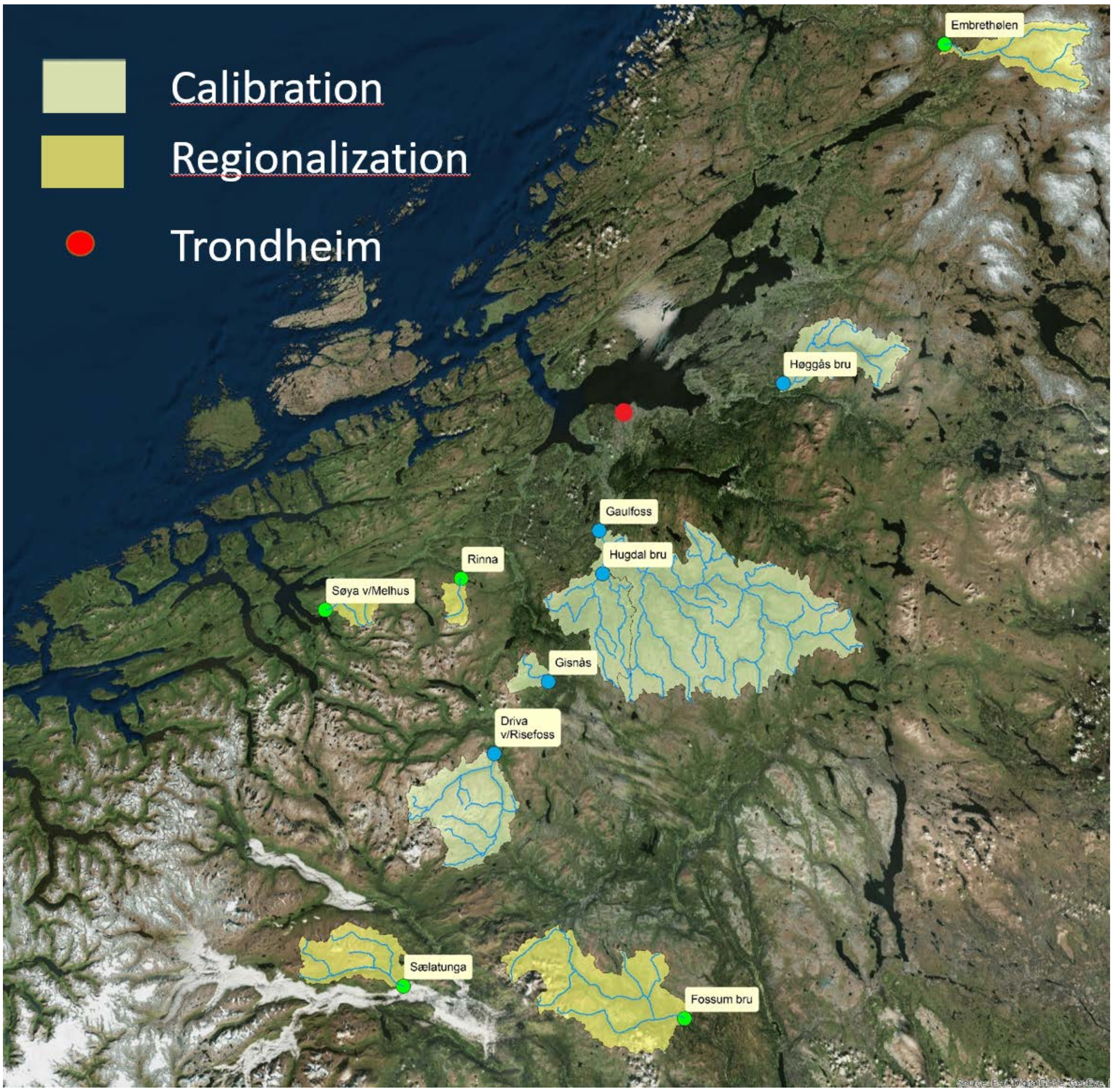

Figure 6 Map of calibration and regionalization catchments in central Norway.

The five test catchments were calibrated with daily time steps for a time span of five years. The model run time for the whole domain is ca. two minutes. The domain is depicted in Figure 6. The figure shows the catchments used for calibration and validation, the position of the gauges is shown as blue or green circle respectively for calibration and validation gauges.

The model's accuracy quantified by the mean KGE has only marginal differences between calibration and validation. As tabulated in Table 5, the KGE of the five gauges ranged between 0.63 and 0.83 for both the calibration and validation period. For the station *Hugdal bru*, there was no discharge data available for the validation period. Judged on the performance based on the KGE of the split sample, the model clearly shows validity.

Table 5 Comparison of KGE of calibration and validation time series for gauges of calibration catchments and proxy catchments.

Station name	KGE Calibration	KGE Validation	Catchment area [km^2]
Driva ved Risefoss	0.73	0.72	746
Gaulfoss	0.83	0.83	3086
Gisnås	0.63	0.67	95
Høggas bru	0.71	0.7	495
Hugdal bru	0.71	Not available	546

Gisnås has the lowest performance and it was also noticeable during the calibration that parameters resulted in performance trade-offs between *Gisnås* and the remaining catchments. The calibration process showed a trade-off between accurate results for small and large catchments.

Model performance in large and small catchments

Overall, large catchments performed better, which was expected because of the following reasons:

- Faulty delineations have less effect on routing and overall water-balance for large catchments because of compensating effects.
- Local effects that can vary on a small spatial and temporal scale, such as radiation, wind and catchment slope were not considered in this model setup. Averaging effects on larger scales weaken impacts of local effects. The impact of this averaging effect is beneficial for the model performance and is proportional to catchment size.
- Small catchments consist of fewer sub-basins, making attenuation and dampening less accurate, since they are calculated once per sub-basin.

An example of the model behaviour is shown with results of the gauge *Driva ved Risefoss* in Figure 7. The following aspects of the modelled time series are explained for this gauge, but do not represent all gauges. For a more detailed description of results, we refer to chapter 4 in Schönfelder (2017).

Peak flows

The flow exceedance curve given in percentiles shows, that the model underestimates peak flows exceeding 99 % of the flow conditions by roughly 25 %. This is further shown in the monthly comparison of observed and simulated runoff in the top right of the graph. The model set-up of south Norway has the same behaviour. It is assumed that the interpolated gridded precipitation data smoothens extreme events to an extent, which leads to a temporal water balance deficit. It may be compensated for in the next model setup by increasing precipitation by a constant percentage.

Low flow

Comparing the hydrographs of simulated and observed discharge showed consistent underestimation of low flow during the summer. This might be based on the pre-existing deficit in water balance of strong precipitation events beforehand, reducing groundwater-fed runoff.

Water balance

There is also a weak tendency to underestimate discharge for the Driva station, as shown in the accumulated volume error curve at the bottom of the figure. This was not the case for all catchments, two simulated gauges have almost no water-balance errors. The average yearly accumulated volume error of the calibration catchments is 5 – 10 % of the average precipitation.

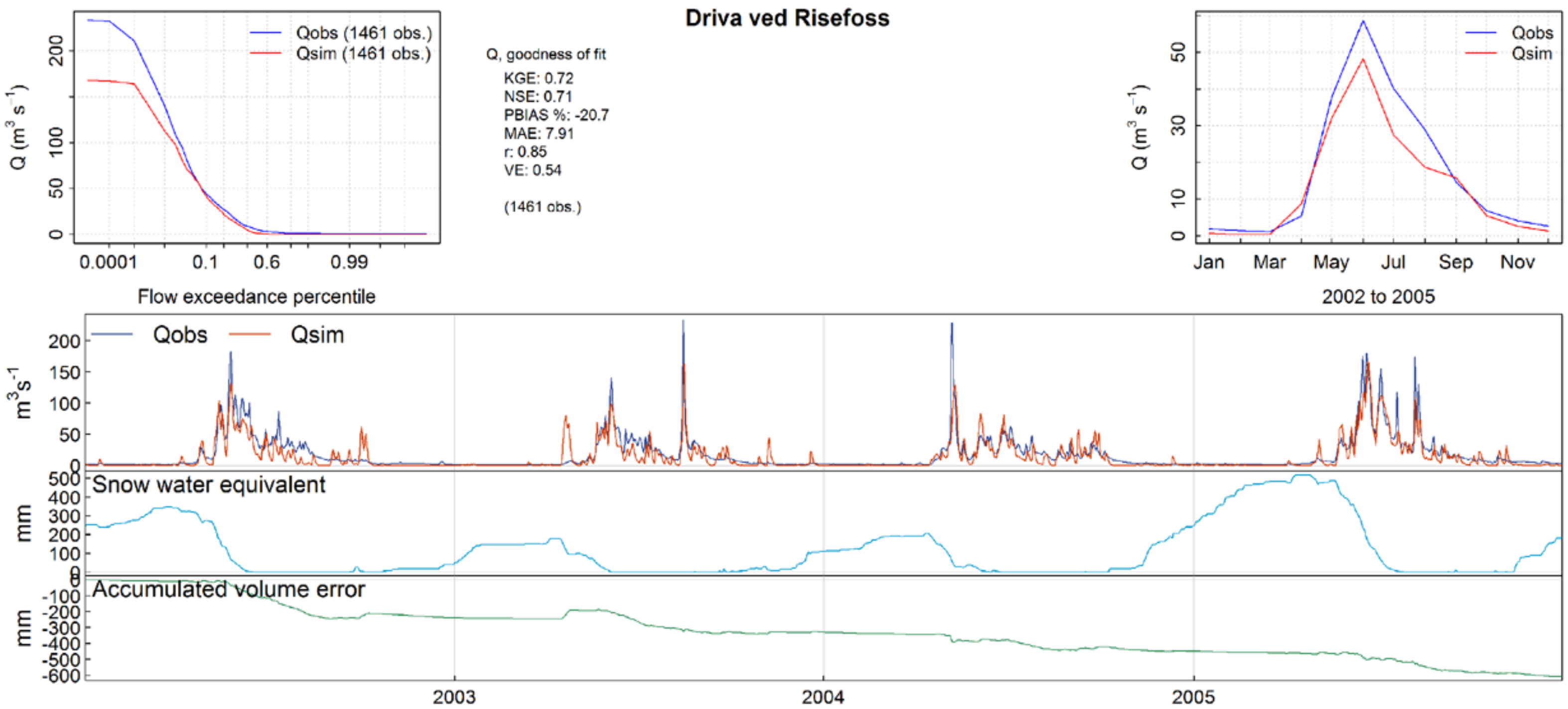

Figure 7 Performance evaluation example of validation period on gauge Driva ved Risefoss.

4.4 Regionalisation of parameter set

The regionalisation, or proxy test respectively, was executed based on the calibration of five gauges for central Norway. The five selected catchments for proxy testing are shown in Figure 6. The selection was focussed on unregulated catchments of varying size within proximity of the calibrated catchments. It was verified that all land uses and soils existing in the proxy catchments were also present in the calibration catchments, to avoid transferring uncalibrated model parameters.

Table 6 List of stations' Performance and size of proxy catchments.

Proxy catchments		
Station name	**KGE**	**Size [km^2]**
Embrethølen	0.50	494
Fossum bru	0.73	1137
Rinna	0.53	88
Sælatunga	0.40	458
Søya ved Melhus	0.30	137

The model results for uncalibrated catchments (tabulated in Table 6) have a high variance and a lower mean KGE than for validation and calibration. *Fossum bru* has the best performance of the proxy catchments. This can assumedly be derived from the size and the land use distributions. *Fossum bru's* catchment is dominated by the land uses "Fjell" and "Forest", both land uses are predominant in all calibration catchments. *Søya ved Melhus* shows the poorest performance of the proxy test cases with the lowest KGE of 0.3 and an accumulated volume error rate of 700 mm/year. Since its catchment also has the highest fraction of "annet areal", it is expected that this land uses' evaporation rate is strongly overestimating evaporation.

The parameter transfer was then tested on 22 other catchments. These catchments were selected based on their near-natural conditions without strong changes due to hydropower regulation. Their catchment delineation and land use similarity was not reviewed. The average KGE for these 22 catchments was 0.51.

Parsimony

The Nevina-based rastermap for land use contains areas that are not classified ("annet areal"), and land use properties for these regions must be calibrated nonetheless. This can lead to an over-parametrized model setup and can lead to imprecise predictions in ungauged catchments. The Corine Land cover map (2012) is complete and distinguishes between more types of land uses.

The soil raster map by NGU is based on the pedogenesis of the soils and not their respective physical / hydrological properties. Their properties are therefore not consistent throughout the domain, meaning that their calibration is not valid for all locations in the domain. The Harmonised World Soil Database (HWSD) has a coarser spatial resolution, but is created based on the hydrological properties such as field capacity and porosity.

The results of the parameter transfer have shown that catchments are prone to poor performance due to high percentages of land uses that are not calibrated well. This is more likely to occur in small catchments. We can therefore assume that a calibration of few land uses and soils lead to overall better results than a complex distribution that may lead to over-parameterization. Also, the transferability of the soil type and land use categories must be ensured. We therefore recommend using the Corine Land Cover map as a database for land uses. For the soil type database, we recommend using the HWSD or a classification that identifies soils based on the occurring land use, since they are associated with each other in many cases.

4.5 Simulated hydrological indices at selected sites

A common approach is to derive hydrological indices from historical discharge time series before and after regulation. In many instances, time series of sufficient length are not available for the period before hydropower regulation. In this work, we calculated the indices for the natural conditions with simulated discharge series from HYPE and compared them with indices derived from historical discharge data under regulated conditions of nine years. Time series from the same periods were used whenever possible, in order to increase comparability between measured and simulated discharge. The main advantage of the model approach is the possible generation of long discharge time series for natural conditions, which is solely dependent on precipitation and temperature time series, which exist for many decades in Norway. Long discharge time series are relevant to define hydrological indices related to return periods and magnitude of floods; the lack thereof can hinder generation of hydrological indices (Acreman et al. 2009).

The selected statistical parameters in this project are calculated from daily discharge time series using an Excel-Spreadsheet created as part of the work of Hohl (2003). The selected indices are based on previous studies that employed IHA (Richter et al. 2006; Hohl 2003; Forseth and Harby 2014). Daily data does not allow for consideration of hydro-peaking related indices, since intraday changes are not covered by the available data. The COSH-tool or other tools to assess and summarize hydropeaking statistics can therefore not be used in this project.

Four test sites in the central Norway domain (see Figure 8), where runoff gauges with available historical discharge time series are bypassed by hydropower diversions, were selected. For the case of *Driva ved Grensehølen*, the upstream section of the river and its tributaries is regulated, the considered station is not directly bypassed. Water is extracted from a big reservoir upstream, used for hydropower production and then released into the fjord. The station *Sjursberget* is directly bypassed and there are diversions from its upstream section. The test cases *Nordsetfoss* and *Brattset* are directly bypassed and their upstream area contains other hydropower plants, but there are no extractions from the catchments.

Information about the stations, the hydropower installation and the catchments are tabulated in Table 7.

Table 7 Regulated river locations used for IHA.

Station name		Nordsetfoss	Driva ved Grensehølen	Sjursberget	Brattset
Subbasin ID		21816	122062	113871	22847
Station ID		123.22.0	109.20.0	112.6.0	121.23.0
River system		Nidelva	Driva	Vinddølelva	Orkla
Start of hydropower operation		1977	1973	1970	1982
Regulated observed time series		1991 - 1999	2002-2010	1976-1984	2008-2016
Unregulated simulated time series		2002-2010			
Power plant name(s)		Bratsberg	Driva, Vassli	Gråsjø	Brattset
Drainage area [km²]	WHIST	3010	1590	167	1460
	Nevina	3006	1630	168	1455
Drainage area difference [%]		0.13	-2.47	-0.82	0.34

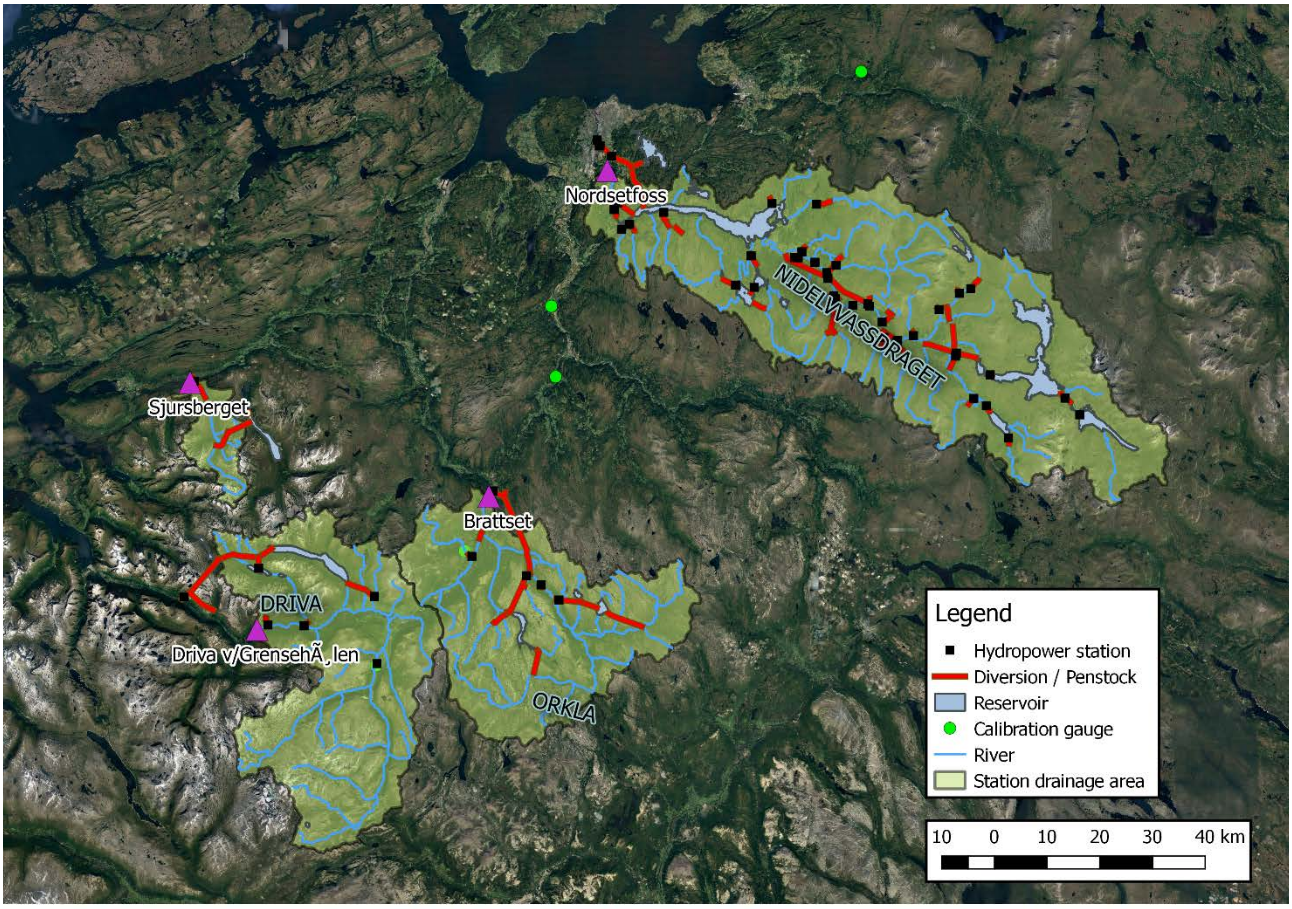

Figure 8 Overview over test sites for calculation of hydrological indices.

The tables Table 8, Table 9, Table 10 and Table 11 show the calculated IHA and the relative deviation from the simulated natural conditions of the sites.

Obs.: The calculated indices are exemplary cases and based on unverified test simulation results. The results have not undergone uncertainty analysis. Their usage as further reference or recommendation is strongly discouraged.

Table 8 Hydrological indices of station Nordsetfoss.

Parameter	Index unregulated	Index regulated	Relative change of hydrological Index
Average runoff [m^3/s]	85.6	42.6	-50 %
Annual 1 day max [m^3/s]	500.3	190.5	-62 %
Annual 30-day max [m^3/s]	274.9	89.8	-67 %
Annual highest 7-day average flow [m^3/s]	417.3	153.4	-63 %
Number of rises [-]	20.9	1.9	-91 %
Number of falls [-]	14.0	1.8	-87 %

Table 9 Hydrological indices of station Brattset.

Parameter	Index unregulated	Index regulated	Relative change of hydrological Index
Average runoff [m^3/s]	31.3	9.8	-69 %
Annual 1 day max [m^3/s]	274.5	143.8	-48 %
Annual 30-day max [m^3/s]	122.4	41.4	-66 %
Annual highest 7-day average flow [m^3/s]	205.2	84.7	-59 %
Number of rises [-]	25.9	6.4	-75 %
Number of falls [-]	19.7	5.0	-75 %

Table 10 Hydrological indices of station Driva ved Grensehølen.

Parameter	Index unregulated	Index regulated	Relative change of hydrological Index
Average runoff [m^3/s]	32.7	28.2	-14 %
Annual 1 day max [m^3/s]	311.9	298.0	-4 %
Annual 30-day max [m^3/s]	134.4	108.5	-19 %
Annual highest 7-day average flow [m^3/s]	241.5	181.0	-25 %
Number of rises [-]	22.0	9.6	-57 %
Number of falls [-]	15.1	7.9	-48 %

Table 11 Hydrological indices of station Sjursberget.

Parameter	Index unregulated	Index regulated	Relative change of hydrological Index
Average runoff [m^3/s]	5.9	5.3	-10 %
Annual 1 day max [m^3/s]	38.8	72.7	+87 %
Annual 30-day max [m^3/s]	20.1	20.2	+0 %
Annual highest 7-day average flow [m^3/s]	30.6	31.7	+3 %
Number of rises [-]	27.7	21.7	-22 %
Number of falls [-]	22.3	20.7	-7 %

Average runoff

The average runoff is changed quite drastically due to hydropower at *Nordsettfoss* and *Brattset* (-50 % and -69 %), both stations are located in the residual flow downstream section of a hydropower plant. For *Driva ved Grensehølen* and *Sjursberget*, this parameter only changes slightly, we assume that the water extraction from their catchment is not significant in comparison to total runoff.

Low-flow related indices

Low flow related indices showed high differences up to ~2000 % between modelled unregulated and measured regulated discharge. As mentioned in sub-section 4.3, the current model setups underestimate low-flow discharges. It is assumed that the difference of this index for all stations is to a large extent based on model inaccuracy. The current model set-up is therefore not able to yield credible results for low-flow indices, they are therefore omitted in this report.

Annual highest 7-day average flow

This parameter is changed drastically for *Nordsetfoss* and *Bratsett* (-63 % and -59 %), the alteration of this index is less drastic for the stations that are not directly bypassed.

Calibrated catchments in the same region showed slight underestimation of periods of extended flooding. This model tendency is also reflected in the comparison of Annual highest 7-day average in the following Method validation (Table 12 and Table 13). It is therefore realistic that the annual highest 7-day average flows and their respective differences to the regulated indices are even higher in reality.

Number of rises and falls

Similar to the change of prior mentioned indices due to hydropower production, the number of rises and falls are extremely reduced for the bypassed stations. The high number for falls and rises of the simulated unregulated discharge is partly due to model behaviour, this aspect is further explained in the following paragraph.

Method validation

The model accuracy was measured on the KGE, which is only one performance metric amongst many. In order to ensure that the calculation of hydrologic indices is not erroneous because of errors in the model results, we compared both indices calculated from model results and from observed discharge without regulation changes.

The indices were computed for the simulated and observed runoff of the stations *Driva ved Risefoss* and *Gaulfoss*, gauges that were used for calibration of the model. For a perfect fit of the calibration (i.e. KGE = 1), there would be no difference in indices. This comparison is tabulated in Table 12 and Table 13. There are minor deviations for the indices between observed and simulated discharge series of both stations for average runoff, annual 1-day-maximum and 30-day-maximum and annual highest 7-day average flow. These deviations are acceptable and within the range of model inaccuracy of a hydrological model of satisfactory performance.

The number of rises and falls show significant deviation, this means that the model produces artificial rises and falls of the hydrograph. The indices calculated from the model results tend to overestimate these numbers. This overestimation is expected to be negatively correlated with the size of the catchment, since station *Driva ved Risefoss* (Drainage area: 746 km^2) has a much higher deviation in Number of rises / falls than *Gaulfoss* (Drainage area: 3086 km^2). This can be partly explained by the local effects described in paragraph "Model performance in large and small catchments" in sub-section 4.3.

Table 12 Indices of observed and simulated discharge of calibration gauge *Driva ved Risefoss*.

Parameter	Index Observed	Index Simulated	Relative change of hydrological Index
Average runoff [m^3/s]	16.5	13.3	-19 %
Annual 1 day max [m^3/s]	174.2	172.6	-1 %
Annual 30 day max [m^3/s]	70.1	68.3	-3 %
Annual highest 7-day average flow [m^3/s]	112.1	128.2	+14 %
Number of rises [-]	12.7	22.0	+74 %
Number of falls [-]	10.8	16.2	+51 %

Table 13 Indices of observed and simulated discharge of calibration gauge *Gaulfoss*.

Parameter	**Index Observed**	**Index Simulated**	**Relative change of hydrological Index**
Average runoff [m^3/s]	80.3	77.7	-3 %
Annual 1 day max [m^3/s]	685.3	652.8	-5 %
Annual 30 day max [m^3/s]	318.2	339.6	+7 %
Annual highest 7-day average flow [m^3/s]	474.2	530.0	+12 %
Number of rises [-]	19.7	26.1	+33 %
Number of falls [-]	16.2	18.8	+16 %

The indices and the validation of the generation method showed that it is useful to calibrate the model for specific model uses. A model calibration for low flows can be set up as well as a calibration that is designated for flood discharges. The different calibrations only make up for a small share of the work involved to set up a large region.

5 Assessment of the suitability of HYPE in Norway

The assessment of the suitability of HYPE for the purpose of supporting the implementation of the EU WFD, was made with use of a set of criteria, inspired by the benchmark criteria developed by Saloranta et al. (2003). These were criteria developed with the need for assessing how appropriate a particular model code is in providing decision support for a specific management task. The criteria set by Saloranta et al. (2003) is not designed to assess the quality of the output, but is assumed to be applied prior to configuring a model. This means that the criteria do not take into account the performance of the model. The criteria and their evaluation results are shown in Table 14.

The selected set of hydrological indices and the change in these indices due to river regulations (before and after regulation) are key outputs relevant for EU WFD, as well as the capability to simulate reference conditions. The suitability of HYPE is assessed with this in mind. The hydrological indices selected are indices mostly related to low flow conditions.

We believe that HYPE must be considered as an expert tool rather than a tool that can be operated by many water managers with limited experience with hydrology and hydrological modelling. As such, less emphasize was put on those criteria important for less experienced users, such as user-friendliness, availability of user documentation and version control. Coming out of the conclusion that HYPE must be considered an expert tool, we have rather emphasized HYPE's ability to generate output which can be made directly available for the regional managers, for instance via Vann-Nett.

Table 14 Systematic approach for the assessment of HYPE's suitability for the specific management of supporting the implementation of the EU WFD in Norway. The assessment criteria are based on the approach proposed by Saloranta et al. (2003).

Criteria	Evaluation results
How well do the model output variables relate to the management task?	Runoff and ground-water level are relevant output variables. Hydrological indices can be calculated based on this output. HYPE is also capable of simulating water temperatures if required calibration data is available, but this was not tested in this study.
Does the model include the key processes relevant to the management task?	Yes, the model is suitable for simulating the conditions in unregulated river basins, and as such capable of generating indices for reference conditions. Parameter values can easily be transferred from calibrated (monitored) basins to ungauged basins. HYPE includes dynamic functionality to simulate systems regulated for hydropower production. The implementation of discharge time series (hourly or daily data) for production flow to of reservoirs makes simulation of effects of operation possible.

Does the model's temporal and spatial span and resolution correspond to the management task?	Temporal resolution and span: Yes, the temporal resolution can be set to hourly or daily for both in- and output. Time span of operation is only limited by availability of precipitation and temperature time series. Spatial resolution and span: Relevant output is only available at sub-basin outlets (runoff) or as sub-basin average (groundwater-level). A fully distributed model that includes cell-to-cell flow could have been better in order to provide results on the spatial scale/extent of a waterbody. The introduction of a fully distributed model might, however, have introduced other challenges). This is explained in full detail in sub-section 3.2 – Delineation of catchments.
Are all the necessary data required for the implementation of the model available?	Yes, there is good correspondence between available data and the data requirements of HYPE, but more detailed input would lead to a more sophisticated model setup and maybe more precise output.
Is this model code potentially suitable for this problem?	*Yes – we believe so.*
Is there sufficient scientific and stakeholder's acceptance of the model code?	Yes, the model is used for the implementation of the WFD in Sweden and set up for many catchments worldwide. Several scientific publications prove the scientific quality of the model.
Is there sufficient guidance to aid model application?	Yes. Extensive documentation of the processes, in/-output and calibration can be found on http://www.smhi.net/hype/wiki/doku.php and https://sourceforge.net/p/hype/discussion/1818967/. Methods for preparation of input data can be found in (Schönfelder 2017). For support with the preparation tool WHIST, there is an active forum: https://sourceforge.net/p/hype/discussion/whist/.
Has the model code been sufficiently tested?	Yes, given the large number of applications with HYPE, the model code has undergone extensive testing and the risk of encountering technical errors hindering future applications in Norway is considered low.
Is the user interface appropriate for the application and user?	In order to modify the model, the user interface is appropriate for experts only. As an example, HYPE does not have a graphical user interface (GUI). The user interface is appropriate for results extractions for users with little HYPE modelling knowledge. A look-up table for the corresponding sub-basins to a point of interest and the editing of a simple text file is sufficient to generate results.

How identifiable are the model parameters?	Yes, the model parameters are to a large extent identifiable because they are linked to physical processes.
Is there sufficient understanding of the model's uncertainty and sensitivity?	We know the sensitivity of the model to selected parameters as explained in Schönfelder (2017). Automatic analysis (Monte Carlo) of sensitivity and uncertainty is possible, but was not thoroughly tested in this study.
Is the model code sufficiently flexible for adaptation, improvements and linking?	Good support is available from SMHI, but adaption and improvements of the model code is difficult to do ourselves. Linking is fairly straight-forward, either by manually taking output and preparing this for the next model, or by scripting in R (or any other tool) and linking automatically output files with HYPE.
Is the model code suitable for this application?	*Yes*

Table 15 Additional criteria for the assessment of HYPE's suitability as a decision support tool for the Norwegian authorities in the implementation of the EU WFD, based on experienced gained with setting up HYPE.

Criteria	Evaluation results
Model performance	HYPE performed well for calibration and validation in both test regions. It performs particularly well for large catchments. The results of different calibration strategies showed that the calibration can be focussed on specific performance objectives, mainly low-flow and peak-flow conditions. The calibration for the south of Norway has its strengths in modelling low-flow conditions and underestimates peak discharges. The calibrated model for central Norway simulates especially the first snow-melt peak of the year well and also underestimates peak flows. Low-flows are underestimated to some extent. The performance, assessed according to the KGE-criteria, is considered well above acceptable. HYPE performs reasonably well in ungauged proxy basins.

6 Conclusions and Recommendations

This study has assessed the suitability of the hydrological model HYPE in supporting the implementation of the EU WFD in Norway with information regarding the hydrological state and changes due to river basin regulations. HYPE was configured for most of the land areas from Nordland and southwards, and based on this we draw conclusions and propose recommendations as given in the following sub-sections.

6.1 Conclusions

From the study, we conclude on the suitability of HYPE:

- HYPE can produce relevant hydrological output from simulations, such as discharge, water temperature and a set of water quality parameters. Based on timeseries of discharge, a set of hydrological indices relevant to assess the ecological state and change can be calculated, such as changes in annual flow values, 7-days minimum flow during winter and summer, and change in return period of floods, if sufficiently long timeseries exist. The calculation of these indices requires a properly calibrated model set-up with respect to those indices to be calculated.

- HYPE has proven to produce results of satisfactory quality, measured as Kling-Gupta-Efficiency (KGE). The cases studies gave an average KGE-performance of 0.73 and 0.72 for the calibrations and 0.73/0.68 for the validations for central and south Norway respectively. Depending on the performance objectives (e.g. low versus high flows), the introduction of two different parameters sets might be needed.

- Model results have demonstrated that it is possible to model large shares of Norway's land surface within a single model and multi-gauge calibration.

- A major benefit of HYPE is to have time series for natural conditions for the past (reference conditions), the present and the future (using land use change scenarios and climate scenarios), even for catchments that have been regulated for decades.

- HYPE is able to simulate discharge in regulated river basins in a temporally fine scale, as there is a routine available to calculate hydropower production and reservoir management explicitly. This function was not evaluated in the current model set-up.

A summary of the advantages and disadvantages of HYPE is shown in Table 16.

Table 16 Summary of advantages and disadvantages of HYPE applied in the context of the EU WFD implementation in Norway.

Advantages	Disadvantages
• Good representation of spatial heterogeneity • Automatic Parameter transfer based on physical properties • Process modules interchangeable and flexible on data basis • Regulation modules are available to dynamically simulate the effects of hydropower (not evaluated in this study) • Supports the calculation of hydrological indices for natural and regulated conditions (regulated conditions not tested in this study) • Further software development & support • Open Source software • Basis for sediment, nutrient and temperature modelling (not tested in this study)	• Tedious and extensive manual preparation of sub-basins and flow network in WHIST • Input data preparation computationally demanding • Expert knowledge necessary to operate the model in order to produce reliable results

6.2 Recommendations for further use of HYPE

For the further use of HYPE in Norway in the context of the EU WFD, we recommend the following concepts for improvement of the simulation results and model utility.

Delineation

The manual corrections of the flow routing, delineation and lake insertion is a tedious and time-consuming task and should be done only once for whole Norway. The delineation should be done in accordance to the national water-body designation, so model results are directly coupled to water-bodies.

Aim-oriented model calibrations

The model setups and their results show that the model performance on individual aspects such as low-flow periods and peak-flows depend on the calibration approach. It is assumed that one calibration (with one parameter set) with good performance on both high- and low-flows is difficult to achieve. We therefore recommend two different model set-ups with their respective focus.

National setup

One of the main features of HYPE is the automatic parameter transfer and the multi-catchment approach. It is therefore meaningful to set-up and calibrate a large domain to achieve a calibration that is based on all types of different catchment properties and therefore more robust in terms of parameter transfer to ungauged catchments.

Databases

It is recommended to use the Corine Land cover map from 2012 and the Harmonised World Soil Database.

Another potential approach is to couple soil properties to land use. Both approaches reduce overall parameter uncertainty. Also, they increase model parsimony and decrease data preparation and calibration effort. Further improvement of model results may be achieved by using a simple model for solar radiation as model input, e.g. based on latitude and day of the year. The model for solar radiation may improve evaporation description.

Use of HYPE for generation of hydrological indices

As we consider HYPE as a tool to be operated by a dedicated expert, we recommend that regional water managers have direct access to the model results, for instance via Vann-Nett, instead of operating HYPE themselves.

7 References

Acreman, M.; Aldrick, J.; Binnie, C.; Black, A.; Cowx, I.; Dawson, H. et al. (2009): Environmental flows from dams. The water framework directive. In *Proceedings of the Institution of Civil Engineers - Engineering Sustainability* 162 (1), pp. 13–22. DOI: 10.1680/ensu.2009.162.1.13.

Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance document no. 13 Overall approach to the classification of ecological status and ecological potential. ISBN: 92-894-6968-4.

Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy (2000). In *Official Journal of the European Communities (OJEC)* (L 327/1).

Forseth, Torbjørn; Harby, Atle (Eds.) (2014): Handbook for environmental design in regulated salmon rivers. NINA, SINTEF, NTNU, UiO, uniMiljø. Nina Special report. Trondheim.

Free Software Foundation (2007): GNU Lesser General Public License, 6/29/2007. Available online at https://www.gnu.org/copyleft/lesser.html.

Gupta, Hoshin V.; Kling, Harald; Yilmaz, Koray K.; Martinez, Guillermo F. (2009): Decomposition of the mean squared error and NSE performance criteria. Implications for improving hydrological modelling. In *Journal of Hydrology* 377 (1-2), pp. 80–91. DOI: 10.1016/j.jhydrol.2009.08.003.

Harby, Atle; Dervo, Børre; Gosselin, Marie-Pierre; Kile, Maia Røst; Lindholm, Markus; Sundt, Håkon; Zinke, Peggy (2017): Metoder for kartlegging, karakterisering og klassifisering av hydromorfologi. Draft SINTEF report.

Hohl, Christian (2003): Effects of a changing climate regime on the runoff in three study catchments. Diploma thesis. NTNU, Trondheim. Department of Hydraulic and Environmental.

Hveding, Vidkunn; Haga, Ingvald; Helland-Hansen, Erik; Holtedahl, Truls; Lye, Kare Arnstein; Hillestad, Knut Ove et al. (1992): Hydropower development. Trondheim: Norwegian Institute of Technology, Dept. of Hydraulic Engineering.

Lindström, Göran; Johansson, Barbro; Persson, Magnus; Gardelin, Marie; Bergström, Sten (1997): Development and test of the distributed HBV-96 hydrological model. In *Journal of Hydrology* 201 (1-4), pp. 272–288. DOI: 10.1016/S0022-1694(97)00041-3.

Nilsson, Christer; Reidy, Catherine A.; Dynesius, Mats; Revenga, Carmen (2005): Fragmentation and flow regulation of the world's large river systems. In *Science (New York, N.Y.)* 308 (5720), pp. 405–408. DOI: 10.1126/science.1107887.

Patil, Sopan D.; Stieglitz, Marc (2015): Comparing spatial and temporal transferability of hydrological model parameters. In *Journal of Hydrology* 525, pp. 409–417. DOI: 10.1016/j.jhydrol.2015.04.003.

Poff, N. LeRoy; Ward, J. V. (1989): Implications of Streamflow Variability and Predictability for Lotic Community Structure. A Regional Analysis of Streamflow Patterns. In *Can. J. Fish. Aquat. Sci.* 46 (10), pp. 1805–1818. DOI: 10.1139/f89-228.

Poff, N. LeRoy; Zimmermann, Julie K. H. (2010): Ecological responses to altered flow regimes. A literature review to inform the science and management of environmental flows. In *Freshw Biol* 55 (1), pp. 194–205. DOI: 10.1111/j.1365-2427.2009.02272.x.

Richter, Brian D.; Baumgartner, Jeffrey V.; Powell, Jennifer; Braun, David P. (1996): A Method for Assessing Hydrologic Alteration within Ecosystems. In *Conservation Biology* 10 (4), pp. 1163–1174. DOI: 10.1046/j.1523-1739.1996.10041163.x.

Richter, Brian D.; Warner, Andrew T.; Meyer, Judy L.; Lutz, Kim (2006): A collaborative and adaptive process for developing environmental flow recommendations. In *River Res. Applic.* 22 (3), pp. 297–318. DOI: 10.1002/rra.892.

Saloranta, Tuomo M.; Kämäri, Juha; Rekolainen, Seppo; Malve, Olli (2003): Benchmark criteria. A tool for selecting appropriate models in the field of water management. In *Environmental management* 32 (3), pp. 322–333.

Schönfelder, Lennart (2017): Performance assessment of the semi-distributed hydrological model HYPE for central Norway. Master thesis. NTNU, Trondheim. IBM. Available online at https://brage.bibsys.no/xmlui/handle/11250/2454731.

SMHI: WHIST. World Hydrological Input Set-up Tool. Version 1.0. Available online at http://www.smhi.se/en/research/research-departments/hydrology/whist-eng-1.22052.

Stanford, Jack A.; Ward, J. V.; Liss, William J.; Frissell, Christopher A.; Williams, Richard N.; Lichatowich, James A.; and Coutant, Charles C. (1996): A general protocol for restoration of regulated rivers. In *Regul. Rivers: Res. Mgmt.* 12 (45), pp. 391–413.

Tallaksen, L. M.; van Lanen, H.A.J. (2004): Hydrological Drought. Processes and Estimation Methods for Streamflow and Groundwater: Elsevier.

Zhu, Zhihua; Chen, Zhihe (2016): Approach for evaluating inundation risks in urban drainage systems. In *Science of The Total Environment* 553, pp. 1–12. DOI: 10.1016/j.scitotenv.2016.02.025.

SINTEF

Technology for a better society

www.sintef.no