HydroCen
NORWEGIAN RESEARCH CENTRE FOR HYDROPOWER TECHNOLOGY

CENTRE FOR ENVIRONMENT-FRIENDLY ENERGY RESEARCH

HydroCen Report nr. 31

NOR-HYPE

A distributed hydrological model for environmental applications in Norway

Lennart Schönfelder, Sacha Baclet, Jessica Sienel, Carolina Saldaña-Espinoza and Ana Adeva-Bustos

NTNU SINTEF NINA

HydroCen

The main objective of HydroCen (Norwegian Research Centre for Hydropower Technology) is to enable the Norwegian hydropower sector to meet complex challenges and exploit new opportunities through innovative technological solutions.

The research areas include:

- Hydropower structures
- Turbine and generators
- Market and services
- Environmental design

The Norwegian University of Science and Technology (NTNU) is the host institution and is the main research partner together with SINTEF Energy Research and the Norwegian Institute for Nature Research (NINA).

HydroCen has about 50 national and international partners from industry, R&D institutes and universities.

HydroCen is a Centre for Environment-friendly Energy Research (FME). The FME scheme is established by the Norwegian Research Council.

The objective of the Research Council of Norway FME-scheme is to establish time-limited research centres, which conduct concentrated, focused and long-term research of high international calibre in order to solve specific challenges in the field.

The FME-centres can be established for a maximum period of eight years HydroCen was established in 2016.

NOR-HYPE

A distributed hydrological model for environmental applications in Norway

Lennart Schönfelder[1]
Sacha Baclet[*]
Jessica Sienel[2*]
Carolina Saldaña-Espinoza[2*]
Ana Adeva-Bustos[1]

[1]SINTEF Energy
[2]University of Stuttgart
*Master Thesis at SINTEF

HydroCen

Schönfelder, L., Baclet, S., Sienel, J., Saldaña-Espinoza, C., Adeva-Bustos, A. 2023.
NOR-HYPE. A distributed hydrological model for environmental applications in Norway.
HydroCen Report 31. SINTEF Energy.

Trondheim, May 2023

ISSN: 2535-5392 (digital publikasjon, Pdf)
ISBN: 978-82-93602-32-3

KVALITETSSIKRET AV
Mauro Carolli

FORSIDEBILDE
Lennart Schönfelder

NØKKELORD
HYPE, distributed hydrological model, Norway, water temperature

KONTAKTOPPLYSNINGER

HydroCen
Vannkraftlaboratoriet, NTNU
Alfred Getz vei 4
Gløshaugen,
Trondheim

www.HydroCen.no

Abstract

Schönfelder, L., Baclet, S., Sienel, J., Saldaña-Espinoza, C., Adeva-Bustos, A. 2023. NOR-HYPE. A distributed hydrological model for environmental applications in Norway. HydroCen Report 30. SINTEF Energy.

This report summarizes the hydrological modelling for environmental applications for the mainland of Norway in the HydroCen work package 3.5. Also incorporated in the report are the findings and results of two master theses: water balances of lakes (chapter 2.8) and time-step testing with different precipitation products (chapter 2.7). The model and scripts for creating new set-ups, pre- and postprocessing are available on Git-hub*.

The precipitation-runoff model HYPE produces relevant hydrological output for the Norwegian mainland. In the HydroCen work package 3.5 study HYPE was set-up with a spatially distributed flow network including lakes and reservoirs. Water management model modules, water temperature, lake water level, and hourly time steps were tested and evaluated.

A single-step calibration approach showed good accuracy with an average KGE (Kling Gupta Efficiency) of 0.56. Predictions in ungauged basins showed promising results.

The current model set-up can be used for hindcasting of natural flow conditions and impacts of hydropower systems on hydrological regimes. It can also be used as a base model for further regional studies. Furthermore, it has potential for future applications for nutrient modelling, climate change impacts, and distributed water temperature modelling.

Lennart Schönfelder, SINTEF Energi, Trondheim, (lennart.schonfelder@sintef.no)

*https://github.com/LennartSchoenfelder/norhype

Sammendrag

Schönfelder, L., Baclet, S., Sienel, J., Saldaña-Espinoza, C., Adeva-Bustos, A. 2023. NOR-HYPE. A distributed hydrological model for environmental applications in Norway. HydroCen Report 30. SINTEF Energy.

Denne rapporten oppsummerer den hydrologiske modelleringen for miljøbaserte analyser og tiltak for fastlands-Norge gjennomført i HydroCens arbeidspakke 3.5. I tillegg inkluderer raporten funnene fra to masteroppgaver: vannstandsmodellering i innsjøer (kapittel 2.8) og kortere tidsskritt med ulike nedbørsprodukter (kapittel 2.7). Modellen og skriptene for å opprette nye oppsett, samt for- og etterbehandling er tilgjengelige på Git-hub*.

Nedbørs-avrenningsmodellen HYPE gir relevante hydrologiske resultater for det norske fastlandet. I denne studien i HydroCens arbeidspakke 3.5 ble HYPE satt opp med et romlig fordelt strømningsnettverk som inkluderer innsjøer og magasin som egne delfelt. Modellmoduler for vannforvaltning og vannkraft, vanntemperatur og vannstand i innsjøer ble testet og vurdert, også med timesoppløsning.

En ett-trinns kalibreringsmetode viste god nøyaktighet med en gjennomsnittlig KGE (Kling Gupta Efficiency) på 0,56. Prognoser i umålte nedbørfelt viste lovende resultater.

Dagens modelloppsett kan brukes til evaluering av hydrologiske påvirkning av vannkraftverk på naturlige strømningsforhold. Den kan også brukes som basismodell for videre regionale studier. Modelloppsettet har videre potensial for fremtidige anvendelser for næringsstoffmodellering, konsekvenser av klimaendringer og distribuert vanntemperaturmodellering.

Lennart Schönfelder, SINTEF Energi, Trondheim, (lennart.schonfelder@sintef.no)

*https://github.com/LennartSchoenfelder/norhype

Content

Foreword

The main purpose of this report is to summarize the work carried out in HydroCen WP 3.5 with the hydrological model (HYPE) for environmental applications for the mainland of Norway. The model hindcasts natural flow conditions for all major water-courses Norway for the past decades. It includes the findings of two master theses for lake modelling and time-step testing with different precipitation products.

We would like to express our gratitude towards our hard-working interns and master students Sacha Baclet, Jessica Sienel and Carolina Saldaña-Espinoza. We would also like to thank Knut Alfredsen at NTNU, Jochen Seidel at University of Stuttgart and Cristian Lussana at the Norwegian meteorological institute for very helpful scientific discussions.

The HydroCen research centre is financed by the Norwegian Research Council, the Norwegian hydropower industry, NVE and the Norwegian environment agency.

Trondheim May 2023, Lennart Schönfelder

1 Introduction

1.1 Why yet another rainfall-runoff model?

Many hydrological model tools are available, some shortcoming that are summarized here leads the motivation to work and set up HYPE hydrological model for Norway.

HYPE stands for HYdrological Predictions for the Environment. It is an open-source, semi-distributed, rainfall-runoff and nutrient transfer model developed by the Swedish Meteorological and Hydrological Institute (SMHI). The model is developed in FORTRAN and is available on Windows and Linux. HYPE is based on the HBV model (Bergstroem, 1975), which is very common for Nordic areas, often for Hydropower production models. There are many versions of HBV available, since the original was developed in 1975 (Lindström et al., 1997).

HYPE has its main advantages in prediction of discharge in ungauged basins, explicitly handling lakes and reservoirs, and water quality modelling. These features are explained in detail in chapter 1.3, and their output - jointly with the distributed model structure - enables the assessment of hydropower impact on rivers in a more detailed way than solely relying on water-balance.

The HYPE 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).

This report focuses on setting-up of the model with a delineation that incorporates lakes, and calibration of the rainfall-runoff features of the model in Norway. In addition, hourly time steps, calibration of lakes and temperature modelling were tested in selected catchments.

1.2 Goal of model set-up

The model set-ups presented in this report aim fulfil the following goals:

Goal 1: Spatial and temporal availability:

The model should be set up and readily available for catchments larger than 100 km² on the Norwegian mainland, particularly in areas where hydrological data is scarce and modelled data is most relevant. Commonly, this applies for undeveloped catchments, bypass reaches, and river reaches downstream hydropower outlets.

The model should be able to provide hourly or daily resolution data and be adaptable to provide higher temporal resolutions as the need arises.

Goal 2: Relevant output spatially distributed and of feasible accuracy

The model should provide good delineation for the entire country of Norway, with a trade-off between finer spatial distribution and increased computational requirements.

Achieving reasonable model performance in different regions, many of which completely ungauged, is not trivial. To reach it, we aim for a universal calibration for a large range of catchment properties. Performance criteria for the model must be chosen respectively and coverage of the area with sufficient station data is a further requirement.

Goal 3: Reasonable performance

The model should demonstrate reasonable performance and validation across a variety of flow ranges, ensuring that it can be used with confidence for environmental management and planning.

Goal 4: Hydrological output for environmental analysis
Commonly useful hydrological variables such as discharge and water-levels of lakes need to be computed by the model. Especially for ecological applications, water temperature and nutrient load can be important variables that can be added and tested or refined in future applications of the model.

1.3 HYPE model

HYPE is a semi distributed precipitation-run off model developed by SMHI, the Swedish meteorological institute. Historically, it was developed to overcome shortcomings of HBV, a common hydrological model used by hydropower producers, authorities, and researchers in Scandinavia.

The catchment areas are divided into sub-catchments, which are linked in a flow network.

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. This also means that hydrological properties of SLC are transferred automatically to uncalibrated catchments, which facilitates predictions in ungauged basins. Many possible process modules in HYPE are similar to processes in HBV, e.g., the day-degree method for evapotranspiration. The concept of SLC’s can be compared to the hydrological response units (HRU) 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 forced by air and soil temperatures.

Lakes and reservoirs that the modellers deem relevant enough (e.g., by surface area or reservoir function) are linked within the flow network as a sub-basin with lake properties. Those lakes are later referred to as *olakes* (outlet lakes). Each lake's outflow is calculated by either an individual or a general rating curve. Lake volume and mean depth can be assigned. Those lakes are called *ilakes* (local lakes). There are several management routines to divert water from *olakes*, to other downstream sub-catchments or inter-basin transfer. Output variables for *olakes* are among others evaporation, outflow and water-level. Lakes and reservoirs below that are not explicitly defined in the flow network are calculated as a special SLC as a fraction of a sub-basin and hence cannot have regulation routines.

Potential evaporation is calculated for each class based on land use and air temperature forcing, other alternative modules are possible, such as Priestly-Taylor and FAO Penman-Monteith.

For a more detailed description of the HYPE model, we refer to the HYPE wiki pages (SMHI, 2022)

2 Methodology

2.1 Model domain

Mainland of Norway covers an area of circa 325 000 km^2 and ranges in latitude from 57 degrees south to 71 degrees north. The elevation ranges from 0 m on the coast and 2469 m above sea level at the summit of its highest mountain.

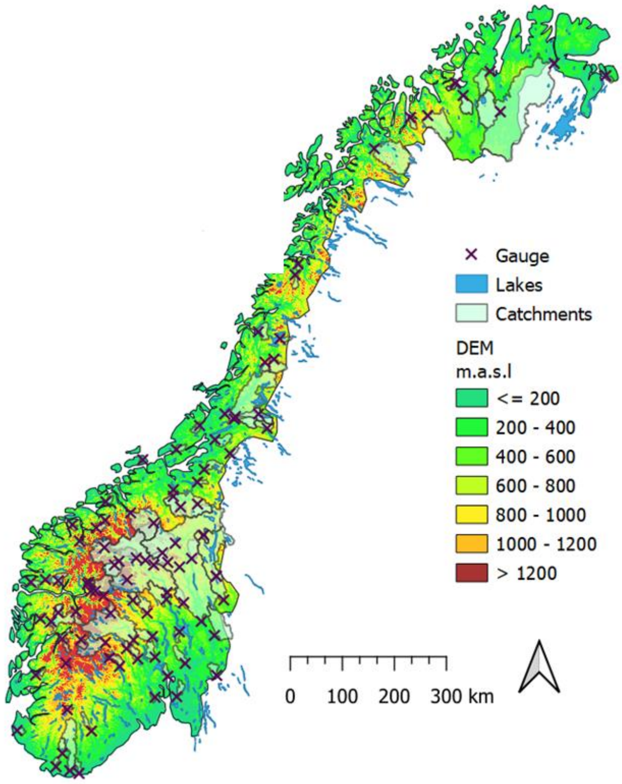


***Figure** 1 Gauging stations, Lakes and catchments used in HYPE set-up for Norway.*

2.2 Forcing and calibration data

For calibration we selected discharge gauges in unregulated catchments. These data have been obtained by filtering stations which have no hydropower regulation in their upstream catchment on NVE Atlas applying the filter *regulerings Type = 100.* Forcing and calibration data for model calibration is summarized in **Table 1**.

Table 1 Data sources for model forcing and calibration.

Variable	Data type	Resolution	timeframe	Source
Daily precipitation	NetCDF	1km x 1km	1957 - 2015	SeNorge2 - (Lussana et al., 2018)
Daily temperature	NetCDF	1km x 1km	1957 - 2015	http://viewfinderpanoramas.org/dem1.html#eur
Daily discharge	csv	Daily average	Varies, min. 5 years	Sildre.no and personal communication with NVE

Daily precipitation and temperature data was averaged using R (R Core Team, 2023) and the ncdf4 package (Pierce, 2023). Some catchments - such as small hydropower reservoirs that are important for routing - do not contain the cell centers of the gridded precipitation data. For these, we interpolate between neighboring grid cell-centers.

We further developed similar software tools to generate precipitation and temperature data generated by the Norwegian climate center from ten different models from the year 1971 to 2100.
(https://klimaservicesenter.no/).

2.3 Delineation and routing

In this sub-chapter, we describe the creation of a flow network of sub-catchments. We focus on automatization of the process and incorporating lakes as sub-catchments.
This is necessary to create the file describing the flow network (*GeoData.txt).*

The delineation process consists in creating polygons representing sub-basins of desired size from a DEM (Digital Elevation Model). We used a DEM with a resolution of 50x50 meters that covers the mainland of Norway excluding Svalbard. Benefit of a resolution this size is that a raster of merged tiles of the entire study site can be held entirely in memory, and all operations can be run without risk of memory overflow errors.

Table 2 Data sources for delineation and routing.

Name	Data type	Resolution	Date of access	Source
DEM Norway	Raster	50m	12.03.2019	
DEM Sweden, Finland, and Russia	Raster	1"	12.03.2019	http://viewfinderpanoramas.org/dem1.html#eur
Lakes, Rivers network, catchments, gauges	Shapefile (polygon, (polyline, points)	-	25.09.2018	https://nedlasting.nve.no/gis/
Lakes Sweden	Shapefile (polygon)	-	17.03.2019	diva-gis.org
Lakes Finland – Ranta 10	Shapefile (polygon)	-	17.03.2019	http://wwwd3.ymparisto.fi/d3/gis_data/spesific/ranta10jarvet.zip

The Swedish and Finnish DEM were merged for the regions where catchments are transboundary (e.g., Trysilelva, Brødbølvassdraget and Nea-Nidelva watershed) in QGIS. All DEM grid cells that were in the sea were distinguished by decreasing the cell values by 1000, the catchment polygons were used to distinguish between land and sea.

Several steps were undertaken to improve delineation quality in comparison to an unconditioned DEM. We used the GIS tool Whitebox GAT (Lindsay, 2016) in combination with python comprising several steps.

1. Walling in of catchment boundaries
Known catchment boundaries were incorporated into the DEM, their borders being a line of higher elevation ("wall-in"). This enables the flow accumulation map and the resulting delineation to be more consistent with known catchment delineations. In comparison to "un-walled" DEM, this reduces misassigning catchment area and therefore minimizes error of total catchment areas.

2. Stream burning
River information was added as lines of lower elevation on the DEM ("burn-in"). River polyline shapefiles were rasterized and then subtracted from the DEM, creating an incision in the DEM where the riverbed is located.

3. Removing single cell pits
Single cell pits are cells with no downslope neighbours, i.e., the cell with highest value of all its adjacent neighbours. They impede overland flow-paths calculated based on flow direction, and therefore have to be removed to allow a continuous connected flow network.

4. Fill depressions
As step 3, this is required to ensure continuous flow from each grid cell to an outlet located along the grid edge.

5. Flow pointer
Steps 1-4 result in a conditioned DEM, from which we calculated a D8 flow pointer grid.

6. Flow accumulation
A flow accumulation raster was created from the D8 flow pointer grid.

7. Sub-basins
Sub-basins were created using the isobasins function with the flow accumulation raster and a target basin size of 25 km2. We selected the isobasin size based on expected computational effort (number of sub-catchments correlates linearly with model wall-time) and required level of spatial distribution.

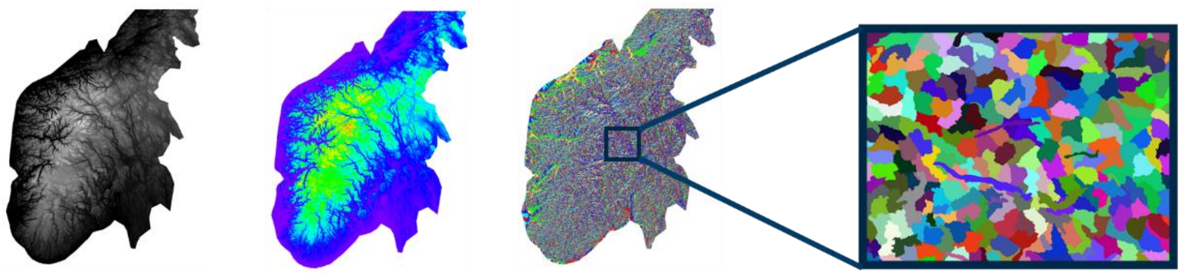


Figure 2 *DEM, conditioned DEM and iso-basins with superimposed lakes.*

8. Lakes and hydropower reservoirs

In this step, lakes with an area > 5 km² will be inserted as sub-basins in the delineation. Islands in those lakes were removed and considered as water. The area error stemming from this assumption will not be associated with the *olake* area, since HYPE reads this area from the correct LakeData.txt, where we used lake areas from polygons. Lakes with an area of less than 5 km² don't have their own catchment and are not explicit in the delineation. They are however considered by the SLC repartition in the sub-basins that they are part of.

The routing process consists in creating a network by linking the created sub-basins together.
Every sub-basin can only flow into one of its neighbouring sub-basins (or the sea). However, several sub-basins can flow into one single sub-basin, in the case of a confluence, for instance. This routing is stored by assigning the ID of the sub-basin that is directly downstream as an attribute of each sub-basin in the shapefile of the delineation. This routing process is automated using Python. The data contained in the resulting shapefile is then saved as a text file that is used as an input for the HYPE model.

For a more detailed description of the delineation steps and information about the employed packages and libraries, we refer to the scripts in the delineation branch of the NOR-HYPE github repository.

2.4 Soil and land use

The model domain consists of a network of sub-basins which are divided into classes. It is possible to divide one basin with different land uses and soil types into several classes. These classes are called "soil type land use combinations (SLC). They can be compared to hydrological response units in HBV. In the Norwegian HYPE model first implemented by Schönfelder (2017) seven SLCs were defined (**Table 3**) The classes are water, mountain, forest, marsh, glaciers and a combination of urban and agriculture land use.

It is possible to model up to three soil layers in one SLC with variable thickness in each layer. Water retention can be varied by the parameters wilting point, field capacity and effective porosity (SMHI 2022). Water retention is allocated evenly between soil layers dependent on their thickness, or these parameters can be defined per soil layer.

Tests showed that a combination of land uses in **Table 4** Soil type land use combinations classes and reclassified soil data from the harmonized world soil database v1.1 (FAO/IIASA/ISRIC/ISS-CAS/JRC, n.d.) resulted in more than 35 SLCs when using 6 soil classes and 7 land use classes, which in turn increases total number of parameters by over 100. Considering model equifinality and the need to reduce it, we chose an approach that reduces numbers of SLC drastically: In this set-up, soil and land use were assumed to be coinciding, meaning that a specific land use has the same soil layers in each region and not a combination of different soils. Although this is questionable from a geophysical standpoint, it reduces the number of possible SLC combinations to five, plus the two types of lakes.

Table 3 *Land use data sources.*

Name	Data type	Resolution	Date of access	Source
Land use Norway, Sweden and Finnland	Raster	100m	12.03.2019	Corine Land Cover (CLC) 2018 (EEA, 2018)
Land use Russia	Raster	250m	19.05.2022	GLC2000 (Bartholomé & Belward, 2005)

Table 4 Soil type land use combinations classes.

SLC	Landuse	Soil	Description	Area fraction [%]
1	1	1	(special class) *Olakes*	5,79
2	1	1	(special class) *Ilakes*	
3	2	2	Mountain	26,66
4	3	3	Forest	35,18
5	4	4	Marsh	26,76
6	5	5	Glacier	0,82
7	6	6	Agricultural + Urban	4,79

Water is divided into two special classes called *ilakes* and *olakes*. An *ilake* is defined as a lake inside a sub-catchment, that can store and evaporate water according to PET. An olake is located at the down-stream end of a sub-catchment, or a whole sub-catchment is defined as an *olake*. Discharge of *olakes* is implemented by using a water level-discharge relationship, or rating curve. Parameters of this relation can be calibrated for the entire set-up (with possibly more than one lake) or added for each lake individ-ually. Another distinction in HYPE is made between main rivers and local streams. Main rivers concen-trate discharge coming from local streams in the sub-catchment. Main rivers also include inflow of water from the upper catchments.

2.5 Calibration strategy

This chapter describes the different calibration strategies that were employed, including calibration data, splitting up the model into different climate zones, available calibration tools and tested performance criteria. It further describes calibration strategies tested to address low-flow conditions.

For the calibration, HYPE provides nine different calibration methods. They are based on either Monte Carlo, Differential Evolution Markov Chain, Brent, method of steepest decent or quasi-Newton (SMHI, 2022). All model set-ups used 2 years of warm-up time to avoid influence of the initial states on the calibration and validation periods.

Most of the model parameters for NOR-HYPE can be calibrated automatically (**Table 12, Appendix**), ex-cept for soil layer depths of the different soil classes, the *icatch* variable and lake rating curve parameters for individual lakes. They were tested during the development of the model, the setup included only a selection of 21 gauged catchments to reduce computational time. The results

For the final calibration we used 102 stations, also shown in **Figure 1** and listed in (**Table 14, Appendix**). For the calibration/validation period we used the periods 2008-2015 and 1982-2007 if not stated other-wise.

2.5.1 Calibration tools

Quasi-Newton methods, which consist in calculating gradients for all parameters and changing their values in the most optimal direction, were determined to be inappropriate for the automatic calibration of dozens of parameters in one calibration run, due to performance issues.

The **Monte Carlo method**, which, in this context, consists in running the model for many sets of random values for all parameters, measuring the performance of the model for these values, and showing the best sets of values at the end of the calibration, gave good results, but was too slow as well.

Two calibration methods showed promising results and are explained in more detail: Differential Evolution Monte Carlo (DE-MC), and Progressive Monte-Carlo method with parameter space limited by best found so far (BP, for bound parameter space).

Differential Evolution Monte Carlo consists in initializing a few independent sets of parameters values with random values (Monte Carlo). These are called a "population" of sets. Those sets are then used in a genetic algorithm: the model is run for all the sets, and the performance of each of these sets is measured using the objective function. The ones that get the best results are then more likely to "mate" or merge with another set of parameters, to create new sets of parameters that will be used in the next "generation" of sets. Some parameter values are also "mutated" or randomly shifted in the process. This process is repeated for a given number of generations. The calibration process thus copies the processes of natural selection and genetic drift in the kingdoms of nature. The specific technique used by HYPE is called "differential evolution".

This method gives very good results, with consistent resulting performance criteria values (both KGE and NSE), although it is comparatively slow. It requires running several tens of thousands of simulations before converging, when many parameters are calibrated at the same time.

BP consists in initializing a given number of independent sets of parameters values with random values (Monte Carlo). The model then runs with all the sets, and the performance of the model is measured for each of them. The values that can be taken by each parameter in the next generation are then bound by the values taken by the top best sets of the current generation. Each new generation is generated with the Monte Carlo method between the newly determined bounds, and this process is repeated for a given number of generations. This method can give very good results and is very fast, but it is not consistent. It requires several hundred simulation runs before converging towards a solution. However, two identical setups with this calibration method can result in significantly different performance criteria values and show high equifinality. In further steps of this research, we decided to use the BP method for the calibration because of its speed advantage and acceptable parametrization with decent goodness-of-fit.

***Table 5** Tested calibration methods and settings.*

Method	Best parameters found	Duration* Number of runs	Time dependency on # parameters	Comments*
Monte-Carlo (MC)	num_mc > 10000	> 10 hours Known a priori	No	Gives good results, when the best sets of parameters are extracted. Slow, but good method to have an idea of the values that the parameters should take (for instance to set bounds for the parameters). BP should be preferred over MC for its higher speed.
Quasi-Newton (Q1 & Q2 & BN)	Not available	Unpredictable Possible to choose a maximum runtime.	Duration increase proportionally with the number of parameters to calibrate	Too slow for the number of parameters calibrated here.
Differential Evolution Monte-Carlo (DE)	DEMC_npop = 1000 DEMC_ngen = 100 Default for the rest: DEMC_gammascale=0.5 DEMC_crossover=0.5 DEMC_sigma=0.5 DEMC_accprob=0	≈ 4 days Fixed from the start by ngen and npop.	No	Effective, but slow method. Yields consistent KGEs when run several times. Enables checking which parameters have the most impact on the simulation results by measuring the spread of the values of the parameters in the last generation . Generates many "close-to-optimal", although different sets of parameter values, in the last generation.
BP	num_bpmc=200 num_bpmax=5 num_ens=5	≈ 2 hours Known a priori	No	Very fast and quite effective. Can reach very high KGEs, but can also converge in a "wrong direction", and thus reach a sub-optimal KGE. Advised to run several of those optimizations parallelly to be sure to get a good result.

*5 years with daily timestep of 200 sub-catchments run on windows on processor Intel(R) Xeon(R) Gold 5122 CPU @ 3.60GHz, 4 Cores, 16GB RAM

2.5.2 Single-step approach

During initial stages of model development, we used a single-step approach for the automatic calibration. All parameters that can be calibrated automatically were were calibrated simultaneously in a 5-year calibration period.

Parameter *icatch* can be manually calibrated for each sub-basin. Due to high computational effort, we decided to calculate it as a global parameter instead.

We used the BP method with num_bpmc=1000, num_bpmax=5 and num_ens=5.

2.5.3 Regionalization

Additionally, to a one-time calibration, a stepwise zonal calibration was tested. This method is also described in Strömbäck et al., 2013, where HYPE was calibrated for Sweden. The main idea is to take one process at a time and calibrate the dependent parameters in each step. In this study, they tried to follow the water flow path from upstream to downstream. A stepwise approach was also done when setting up the World-Wide-HYPE model (Arheimer et al., 2020). They also divided the catchments into different climate zones. There, eleven processes were defined and calibrated iteratively.

To avoid overcompensating errors with different processes, the initial calibration (with all parameters) was set up to already have a high number of simulations, to provide stable results foreach parameter set. That way, the stepwise calibration should just lead to an improvement of parameter sets (fine tuning) instead of changing the model dynamic of the whole catchment. In the final version, the following processes and their parameters were calibrated:

- **Soil:** wcfc1, wcfc2, wcfc3, wcwp1, wcwp2, wcwp3, wcep1, wcep2, wcep3, mperc1, mperc2, sfrost, rrcs1, rrcs2, rrcs3, srrate, macrate, mactrinf, mactrsm, frost, srrcs, surfmem, ttrig, treda, tredb, depthrel
- **Snow:** ttpd, ttpi, deepmen, cmlt, ttmp, fscmax, fsck1, fsceff, fscdistmax, fscdist0, fscdist1, sdnsnew, snowdensdt, fsclim, fsckexp pcusnow
- **Divers:** lp, epotdist, cevp, rivvel, damp,deadl, deadm, pcurain, icatch
- **Olakes(general):** gratk, gratp, grata
- **waterT:** tcfriver, tcflake, scfriver, scflake ccfriver,ccflake, lcfriver, lcflake, t2trlake, t2trriver, t2mix

Note that *olakes* parameters are general for all *olakes* that don't have individual rating curves assigned.

The model calibration was tested by splitting Norway into different climate regions according to a calibration strategy that was applied for HBV (Huang et al., 2019). We used the same zones as Huang et. al 2019 defined. For catchments that overlapped with several zones, we assigned the zone by largest share of areal overlap. Main advantages of the regionalization is parallelising model runs and potentially improved model performance for individual climate zones, without losing the ability to predict in ungauged basins.

Table 6 *Climate zone distribution of model sub-catchments.*

Region	Stations	Sub-basins
Climatezone 1	3	726
Climatezone 2	36	838
Climatezone 3	19	272
Climatezone 4	23	2076

Climatezone 5	16	580
All calibration catchments	97	3828
All Norway	97	16581

2.6 Performance objectives

Choice of performance criteria

The performance criteria available in HYPE are numerous. They include the average NSE (Nash-Sutcliffe efficiency), the median NSE, median KGE (Kling-Gupta Efficiency) and the average of the Pearson correlation coefficient for all basins with observations.

KGE was chosen as metric to assess the performance of the simulation of water flow compared to observations, because it is less biased, unlike the often-used NSE. Moreover, a high KGE usually implies a high NSE, while it is not the case the other way around (Gupta et al., 2009). The KGE has three contributing terms:

$$KGE = 1 - \sqrt{(r-1)^2 + \left(\frac{\sigma_{sim}}{\sigma_{obs}} - 1\right)^2 + \left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2}$$

with r as correlation coefficient between observations and simulations, σ being the variance, and μ being the mean flow to account for bias.

$$NSE = 1 - \frac{\sum_{t=1}^{T}(Q_{sim}^{t} - Q_{obs}^{t})}{\sum_{t=1}^{T}(Q_{sim}^{t} - \overline{Q_{obs}})}$$

Where Q_{sim}^{t}corresponds to simulated discharge and Q_{obs}^{t} to observed discharge of a timestep t, and Q_{obs} corresponds to mean observed discharge.

Establishing what extent of delineation error is acceptable by applying methods shown in chapter 2.3 plays a role within decision making in refining delineation. One theoretical approach focussing on the water-balance is shown in **Error! Reference source not found.**.We assume uniform precipitation and otherwise neglecting potential effects of under- or overestimation of sub-catchment sizes on the model's behaviour. Introducing this time-independent bias to NSE and KGE results in a linear deterioration in KGE caused by an introduction of bias in the term $\left(\frac{\mu_{sim}}{\mu_{obs}} - 1\right)^2$ and a parabolic deterioration in NSE. Any deviation from perfect water balance introduced by erroneous delineation results in a stronger decline in absolute value of performance metrics.

It is important to note here that NSE and KGE scores should otherwise not be compared directly, since they describe good-of-fitness on different scales. A simple comparison can be done by using the mean flow as predictor: this would result in NSE = 0 and KGE ≈ -0.41 (Knoben et al., 2019). Even though both metrics have 1 as their maximum, the user's intuition and subsequent assessment of goodness-of-fit about NSE should not be applied to KGE.

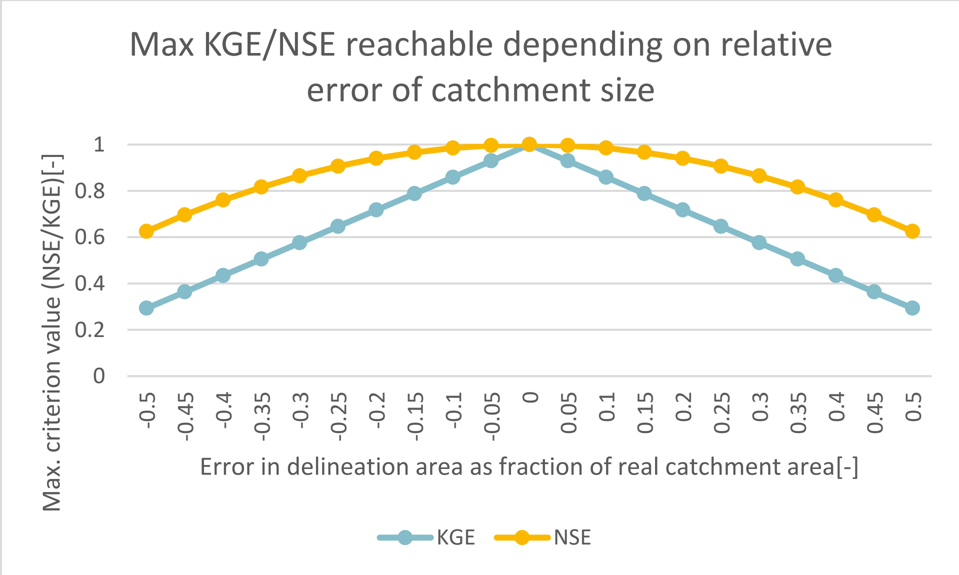


Figure 3 *Theoretical performance scores depending on catchment area error.*

Low flow focus

The prior NOR-HYPE models tend to underestimate low flows (Schönfelder, 2017). Testing during the development phase with reducing the observation dataset to Q90 and below showed that performance To get a good overview of the performance of a model calibration in low-flow conditions, the following strategies were employed. Santos et al., 2018 showed that KGE should not be used together with log-transformed flow series, so we used NSE for the following approaches:

log(cout+1)

The logarithm of the output discharge (cout) + 1 was used as the output variable of the simulation for the calibration with NSE. It slightly improves the accuracy of the calibration for medium-flow (not for low-flow), but noticeably impacts the accuracy for high-flow.

log(cout+0.1)

The logarithm of the output discharge (cout) + 0.1 was used as the output variable of the simulation by changing the code and recompiling of HYPE for the calibration. It noticeably improves the accuracy of the calibration for medium flow and a bit for low flow) but decreases accuracy for high flow. It can be noted that when defining the output variable as log(cout+x), with x>0, the higher x, the lower the importance of low flow in the performance criterion value.

2.7 Time step and precipitation data testing

Jessica Sienel tested HYPE with hourly timestep and using different precipitation products in her Master Thesis (Sienel, 2022). In order to select the catchment for the study, those that were influenced by hydropower regulation were excluded. The test case sites were selected based on the amount of precipitation gauges available in the catchment, proximity of radar stations, variety of hydrological regimes and model uses in the context of other HydroCen research. Nausta and Surna catchments were partly selected for its relevance in the DynaVann project (HydroCen) about the link between low-flow conditions in winter and egg survival of salmon (https://www.ntnu.no/hydrocen/dynavann). Grunnåi was partly selected because of its relevance in the TwinLab project (https://www.ntnu.no/hydrocen/twinlab), where we couple the HYPE project with production models and hydraulic models. For a more detailed description of the test cases, we refer to Sienel, 2022.
The used model-setup was adjusted from what is described in chapters 2.1 to 2.6 of this report from daily to hourly time-steps.

To have a representative study over different regions in Norway the catchments were compared to a study of Gottschalk et al., 1979, that divided Norway into different hydrological regions. According to this study, Norway consists of six regimes. One of them being a mountain regime (H1L1), with high discharges in spring caused by snowmelt and low discharges in winter because of snow accumulation Another regime is called Atlantic regime (H3L3) which is more influenced by rainwater instead of snowmelt and has low flows during summer caused by less precipitation and a higher evapotranspiration. The inland regime (H2L1) is a transition zone between dominant rain and snowmelt discharges with low discharges in winter. There are three other regimes, two of them being transition regimes and one is the Baltic regime (H2L3) with a mix of rain and snowmelt water in spring and low flows in summer. The categorization of the test catchments in the thesis is summarized in **Table 7**.

Table 7 *Location of the catchments according to their hydrological region (Gottschalk et al. 1979).*

Regime	Catchments
Mountain regime (H1L1)	Gaula, Usma, Surna, upper basin of Naust
Inland regime (H2L1)	Grunnåi
Atlantic regime (H3L3)	Lower basin of Nausta
Baltic regime (H2L3)	Parts of Gaula and Surna
Transition regimes (H2L2)	-
Transition regimes (H3L2)	-

Precipitations for these catchments were derived from SeNorge2 and NWP. SeNorge2 is available between 2010 and 2017, NWP between 09/2013 and 2022. To avoid accumulated snow at the beginning of the calibration or validation period, a hydrological year was defined to start in September. The two datasets overlap between 09/2013 and 2017. To have an initial period of one year (e.g., to fill soil water storages), the compared time slot was defined between 09/2014 and 09/2016. To get the best possible results in this interval, calibration was done at that time. The calibration period was chosen to be three years, validation two years. This leads to the calibration of seNorge2 being between 09/2013 and 09/2016 and NWP between 09/2014 and 09/2017. Discharge time series were available between 2010 and 2022 for most catchments except for Usma (2013-2017) where validation was not performed.

HYPE does currently not support calculating PET using Priestly-Taylor or FAO Penman-Monteith formula for hourly timestep. The available method for hourly timesteps is the default model. This method assumes, that PET (*epot*) is dependent on air temperature, with no evaporation below a threshold value (*ttmp*). Furthermore, PET is dependent on land use which is included with the rate parameter (*cevp*). To also account for seasonal changes, a parameter *cseason* is added to the equation. If necessary, a regional correction factor (*cevpcorr*) can be added.

$$epot = (cevp \times cseason) \times (T - ttmp) \times (1 + cevpcorr)$$

Cevp and *ttmp* can be calibrated with automatic calibration. Preliminary results showed that PET was not similar for each calibration. This may be caused by low sensitivity of the model towards evapotranspiration changes. Additionally, a diurnal cycle is not considered, which means that on warm nights, the evapotranspiration was still high, although no sun radiation was available. To get a more physical and not calibration-based PET, it was calculated separately and added to HYPE as additional input data. The method of (Hargreaves & Samani, 1985) was applied for that purpose. This is a simple approach using the minimum (T_{min}) and maximum (T_{max}) temperature of a day to calculate the daily

$$epot = 0.0023 \times R_A \times (T_{max} - T_{min})^{0.5} \times \left(\frac{(T_{max} - T_{min})}{2} \times 17.8\right)$$

Daily extra-terrestrial radiation (RA) can be calculated according to FAO guidelines (Allen, 1998) by only using the latitude (lat) of the area and the day of the year (J). These two values are necessary to calculate latitude (ϕ) in rad, solar declination (δ) and sunset hour angle (ωs).

$$= \frac{24}{\pi} \times 4.92 \times (\omega_s \times sin(\varphi) + cos(\varphi) \times cos(\delta) \times sin(\omega_s))$$

$with\ \varphi = \frac{\pi}{180} \times\ lat$

$$\delta = 0.409 \times sin\left(\frac{2\pi}{365} \times J - 1.39\right)$$

$$\omega_s = arccos(-tan(\varphi) \times tan(\delta))$$

Based on Hargreaves & Allen (2003) and Zarei et al. (2015), Hargreaves method shows acceptable results compared to Penman-Monteith, with less data input necessary. In Sienel (2022), evapotranspiration was calculated by using seNorge2018 daily minimum and maximum temperature data, because it was the most recent product that is available for the whole time period of this study. A weighting function was used to transform daily evapotranspiration into hourly information, having the highest values in the middle of the day and no PET at night (**Figure 4**).

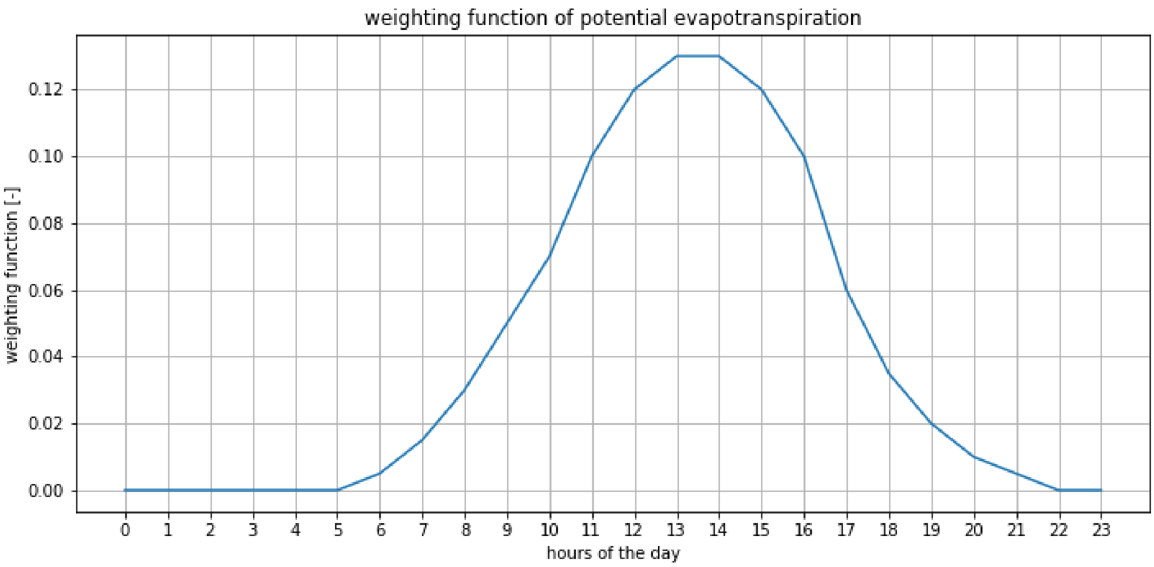


Figure 4 *Weighting function used for calculating the hourly potential evapotranspiration.*

Seasonality was also considered since Norway's day and night-time varies strongly between summer and winter. However, evapotranspiration was found not to be a strong influence on the timing of the model and is more important for the general water balance than for the hourly values. That is why only one weighting function was applied for summer- and wintertime.

PET calculated with Hargreaves, MOD16 PET and AET data were compared. This was done by plotting a timeseries of the evapotranspiration products. Additionally, the annual actual evapotranspiration amount from MOD16 and modelled by HYPE was evaluated (**Figure 5**).

Results shows that HYPE Hargreaves values are typically lower than PET and exceeding AET, meaning that the calculated PET is lower than the one given from MOD16. Still, it is never lower than AET. There are no values of MOD16 in the winter months December and January, probably because it is assumed to be no evapotranspiration then. Hargreaves is close to zero in these months.

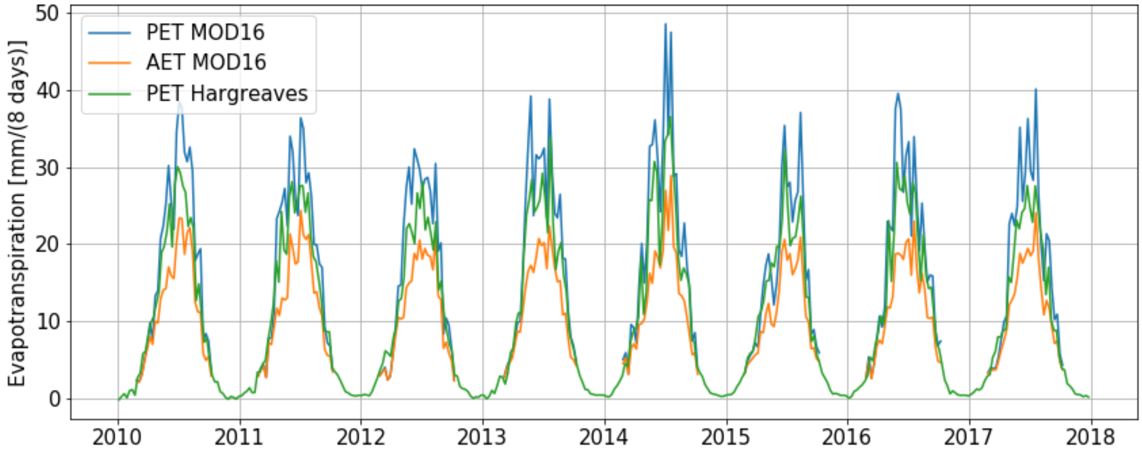


Figure 5 *Evapotranspiration from Hargreaves and MOD16 over time (2000-2018) of the Gaula catchment*

2.8 Lake module calibration

We tested a selection of 21 catchments in the NOR-HYPE setup (chapter 2.1) and divided them into two groups: one “lake-group” of 8 catchments which contained large *olakes* and a “no-lake-group” consisting of 13 catchments without large lakes, but possibly containing smaller lakes considered as *ilakes*. Lake depth information was set to 10m, which is justified by water-level variation of natural Norwegian typically being much smaller than 10m. The average KGE in the lake-group was 0.07 lower than in the no-lake-group.

We then calibrated all parameters of 21 catchments in a 2-step approach: first we calibrated all parameters but the lake parameters, in the second step we only calibrated general lake parameters for all lakes simultaneously. This yielded almost identical performance metrics for both groups.

Carolina Saldana’s Master thesis (Saldana Espinoza, 2022) aims to develop model strategies for improving lake dynamic modelling with natural flow conditions for discharge and water level in HYPE. Seven Norwegian lakes were modelled and calibrated trough stepwise approach. Arheimer and Lindström (2013) considered a stepwise calibration approach with five main steps. The first consisted of calibrating general parameters for defining the processes of evaporation, snow density, and precipitation. The next steps consisted of the calibration of the soil parameters for the definition of infiltration, percolation and surface runoff. The next step was calibrating the land use parameters. The final steps were the river and lake parameters calibration. In addition, Adera et al. (2018) has applied a stepwise calibration for the south of Norway, considering general parameters, soil parameters, land use parameters, river and lake parameters with satisfactory results. In addition, Lindström (Lindström, 2016) denotes that calibration for lakes in HYPE is difficult due to the interdependence of the water reference and rate. For water level calibration, the water reference can be adjusted by the bias that is calculated by the rest of the simulated water level reference minus the observed water level without the water level reference.

Saldana Espinoza (2022) developed two calibration strategies:

Strategy A based on six steps (**Figure 6**) distributed in three stages, which combine manual and automatic calibration. The first stage represents the discharge model, the second is the water level model, and the third is a simultaneously model for discharge and water level. This strategy focuses on the calibration of the model using only lake parameters.

Step	Calibration type	Method	Specification	Forcing data
First stage: HYPE model for discharge prediction without specific rating curve coefficients				
1	Automatic	Progressive Monte-Carlo method with parameter space limited by best found so far (10 000 simulations per step)	Calibration with "**best parameters**"	Discharge
2			Calibration with **lake parameters**	
Second stage: HYPE model for water level prediction with specific rating curve				
3	Manual	Initial run with rating curve approx. from observed data → bias correction of reference water level	(*clwc* and *clws*)<- criterium for KGE (*clwc* and *wstr*) <- bias	Water level
4		Introduction of new water reference w'_o	Corrected w'_o from 3rd step	
5		Manual variation of the lake parameters *k* and *p*	Changing *k* and *p* New *k'* and *p'*	
Third stage: HYPE model for discharge and water level prediction with specific rating curve coefficients				
6	Automatic	Progressive Monte-Carlo method with parameter space limited by best found so far (10 000 simulations per step)	Calibration with "**best parameters**" and new coefficients of rating curve w'_o, *k'* and *p'*	Discharge

Figure 6 *Lake calibration strategy A*

Step	Calibration type	Method	Specification	Forcing data
First stage: HYPE model for discharge prediction without specific rating curve coefficients				
1	Automatic	Progressive Monte-Carlo method with parameter space limited by best found so far (10 000 simulations per step)	Calibration with "**best parameters**"	Discharge
2			Calibration with **soil parameters**	
3			Calibration with **snow parameters**	
4			Calibration with **other processes parameters**	
5			Calibration with **lake parameters**	
Second stage: HYPE model for water level prediction with specific rating curve				
6	Manual	Initial run with rating curve approx. from observed data → bias correction of reference water level	(*clwc* and *clws*)<- criterium for KGE (*clwc* and *wstr*) <- bias	Water level
7		Introduction of new water level reference w'_o	Corrected w'_o from 6th step	
8		Manual variation of the lake parameters *k* and *p*	Changing *k* and *p* New *k'* and *p'*	
Third stage: HYPE model for discharge and water level prediction with specific rating curve coefficients				
9	Automatic	Progressive Monte-Carlo method with parameter space limited by best found so far (10 000 simulations per step)	Calibration with "**best parameters**" and new coefficients of rating curve w'_o, *k'* and *p'*	Discharge
10			Calibration with **soil parameters** and new coefficients of rating curve w'_o, *k'* and *p'*	
11			Calibration with **snow parameters** and new coefficients of rating curve w'_o, *k'* and *p'*	
12			Calibration with **other processes parameters** and new coefficients of rating curve w'_o, *k'* and p'	

Figure 7 *Lake calibration strategy B*

Strategy B This approach is based on 12 steps distributed in three stages (Error! Reference source not found.) which combine manual and automatic calibration. The result is the creation of three models. The first two are independent models for discharge and water level, and the last one models discharge and water level simultaneously. The difference with strategy A is that in the first and third stages, the automatic calibration is performed per specific process.

2.9 Water temperature testing

The model's ability to predict water temperature in HYPE was tested in the Gaula catchment.

In addition to parameter groups for discharge prediction, we added a final group of water temperature related variables and employed the step-wise calibration approach described in chapter 2.5.3.

We used a weighted performance criterion approach with 50% mean KGE of all discharge stations shown in figure **Figure 8** and 50% mean NSE of simulated water temperature in Eggafoss and Hågå bru. Calibration period was 2012-2018 with all calibration time series fully available.

In order to account for hydropower regulation, we first set up a specific delineation where all tunnels starts, and ends coincided with sub-catchment outlets. Water diversion was accounted for in the model file *BranchData.txt*. Required tunnel capacities were retrieved from openly accessible concession documents of TrønderEnergi (*Revisjonsdokument for Reguleringer i Lundesokna Og Sokna*, 2019).

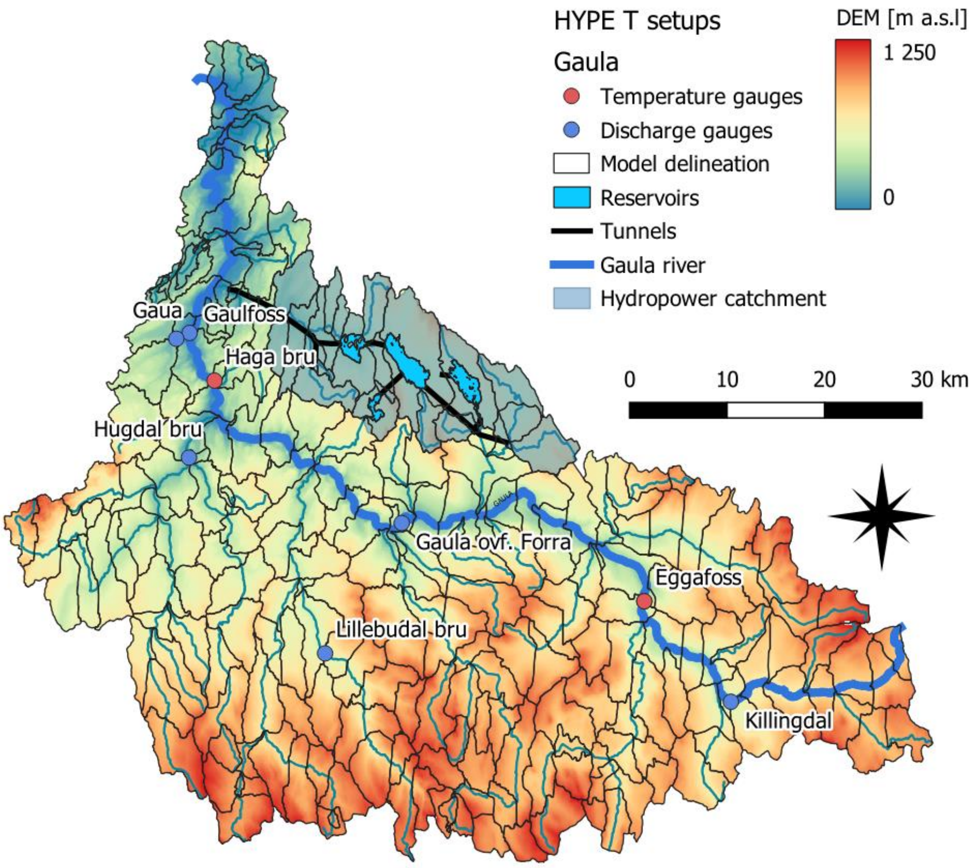


Figure 8 *Gaula – distributed water temperature model setup*

3 Results

3.1 Results distributed model

3.1.1 Results for manual calibrations

The stream depth was calibrated with one value for all SLCs and with a total soil depth of 2m (**Table 8**). Local minimum for model performance can be detected at 1m depth, and optimum for 2m, with a total range of 0.022 KGE difference.

Table 8 *Performance*

Stream depth	Average KGE
0.5	0.828
1.0	0.797
1.5	0.842
1.75	0.843
2.0	0.850

3.1.2 Whole Norway setup

The single-step approach yielded a median KGE of 0.569 and 0.583 for calibration and validation respectively. The distribution of catchment KGE is shown in **Figure 9** and for validation mapped in **Figure 10**

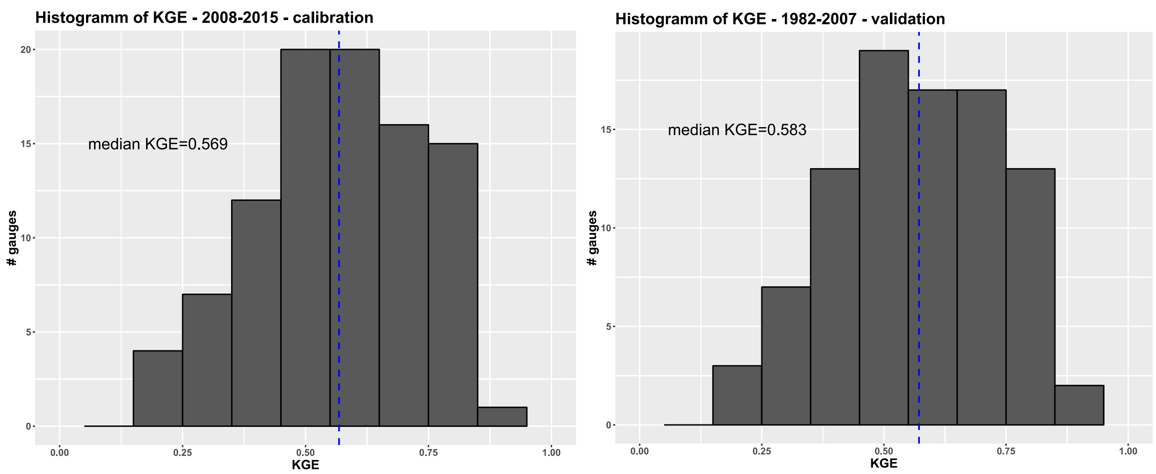


Figure 9 *Model performance for 95 stations – calibration on the left, validation on the right.*

An overview of all stations can be found in **Table 14** in the appendix.

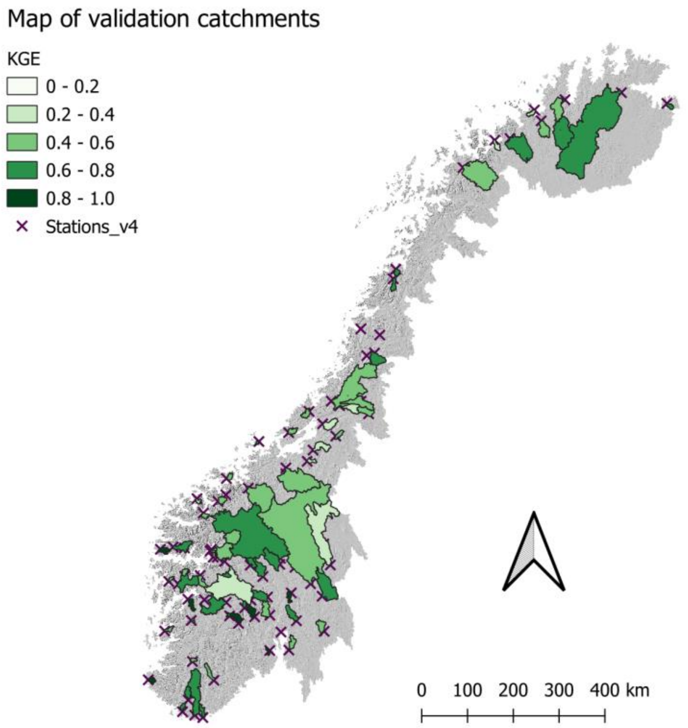


Figure 10 Map of catchment KGE Validation period

Low flow results:

log(cout+1)
The log transformed discharges slightly improves the accuracy of the calibration for medium-flow but not for low-flows, but noticeably impacts the accuracy for high-flow negatively.

log(cout+0.1)
This approach noticeably improved the accuracy of the calibration for medium flow and a bit for low flow but decreases accuracy for high flow. It can be noted that when defining the output variable as log(cout+x), with x>0, the higher x, the lower the importance of low flow in the performance criterion value.

In a majority of catchments, the model underestimates low flows significantly. **Figure 11** shows this. The KGE reaches reasonable model performance only when observations above the 70th percentile are included. Below that, i.e., when only the lowest 20% of observation are used to calculate performance criteria, performance drops very low.

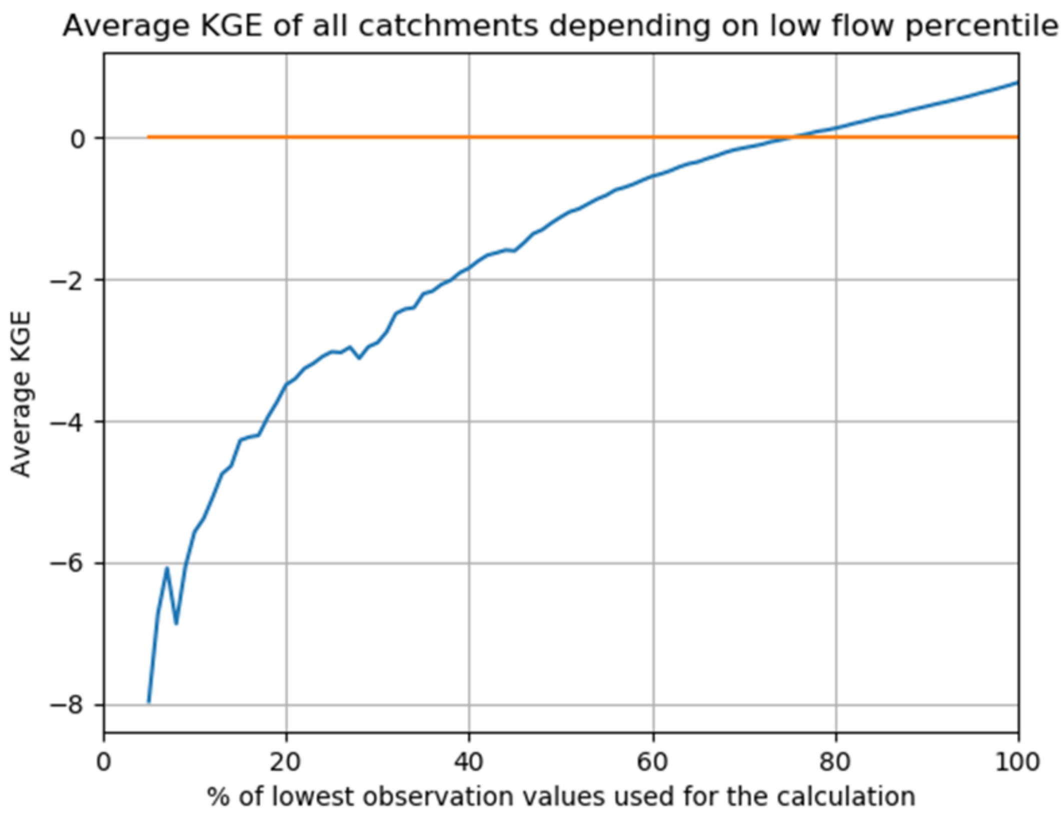


Figure 11. Relationship between model performance and included flow percentiles at Haga bru station

3.1.3 Climate zone calibration results

At the time of writing of this report, the zones 3, 4 and 5 were not finished calibrating (**Table 9**). Model performance is higher than other approaches in climate zone 1, but lower in climate zone 2.

Table 9 Climate zone calibration - KGE overview

Region	Stations	Average KGE
Climatezone 1	3	0.66
Climatezone 2	36	0.5
Climatezone 3	19	-
Climatezone 4	23	-
Climatezone 5	16	-

3.2 Lake modelling results

Table 10 compares the results per stage for calibration strategies A and B. In the first stage, models that simulate only discharge show an average increase of 1.36% in KGE compared to the models that simulate both discharge and water level in the third stage, for strategy A. For strategy B, this increase is 0.67%. Meanwhile, water level models in the second stage led to 1.23% and 0.35% higher KGEs than the model generated in the third stage that combines discharge and water level, for both strategies. Calibration strategy B yields the highest KGEs for discharge, with an average KGE of 0.81 and 0.80 in the first and third stages, respectively, which is higher than calibration strategy A with 0.78 and 0.77, respectively.

The median KGE for discharge in the first and third stages of calibration strategy B are higher than the respective stages of strategy A. However, strategy A yields the highest KGEs for water levels, with an average value of 0.90 and 0.89 in the first and third stages, respectively, which is slightly higher than strategy B with a value of 0.89. It demonstrates that the median KGEs for water level in the first and third stage in calibration strategy A differs by 0.05, while in strategy B, the difference is marginal. Additionally, it is worth noting that the KGE of the model for water level is greater than the KGE for flows. For example, for calibration strategy A, the KGE difference in average between the first step and second step is 15.4% more, while for calibration strategy B, it is 10.7%.

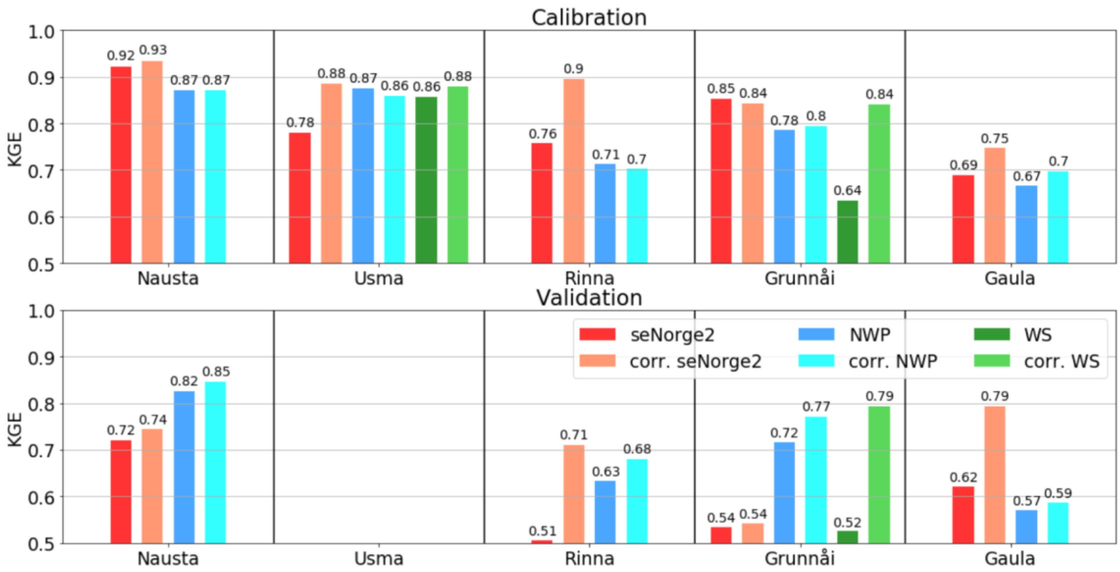
***Table 10** Lake modelling results KGE with strategies A and B.*

Strategies		Calibration strategy A				Calibration strategy B			
Lake	Period	Stage 1	Stage 2	Stage 3		Stage 1	Stage 2	Stage 3	
		Discharge	Water level	Discharge	Water level	Discharge	Water level	Discharge	Water level
Fustvatnet	Calibration	0.81	0.94	0.82	0.93	0.86	0.95	0.86	0.92
	Validation	0.83	0.93	0.84	0.93	0.87	0.94	0.87	0.93
Fijhpelogkoe	Calibration	0.84	0.94	0.84	0.94	0.87	0.95	0.87	0.94
	Validation	0.72	0.94	0.72	0.86	0.75	0.87	0.75	0.85
Gjende	Calibration	0.81	0.95	0.82	0.94	0.83	0.95	0.83	0.94
	Validation	0.64	0.90	0.65	0.90	0.65	0.90	0.66	0.90
Engeren	Calibration	0.92	0.92	0.94	0.95	0.93	0.94	0.95	0.94
	Validation	0.73	0.83	0.76	0.86	0.69	0.75	0.76	0.84
Hestadfjorden	Calibration	0.88				0.93			
	Validation	0.82				0.86			
Vangsvatnet	Calibration	0.92	0.93	0.90	0.87	0.96	0.93	0.95	0.92
	Validation	0.85	0.89	0.84	0.85	0.88	0.88	0.88	0.88
Furusjøen	Calibration	0.58	0.83	0.53	0.82	0.60	0.83	0.59	0.80
	Validation	0.59	0.81	0.58	0.82	0.62	0.81	0.64	0.80
Average		**0.78**	**0.90**	**0.77**	**0.89**	**0.81**	**0.89**	**0.80**	**0.89**

3.3 Hourly time step results

A model performance overview is given in **Figure 12**. Validation of Usma was not possible due to very limited availability of time series.

***Figure 12** Calibration and validation KGE of different precipitation products*



Regarding the performance of seNorge2 and NWP precipitation data in calibration, it was found that seNorge2 yielded higher Kling-Gupta Efficiency (KGE) values than NWP data. However, validation results showed lower performance for seNorge2. Incorporating a correction factor for seNorge2 resulted in further improvement in model performance, suggesting that the data may suffer from measurement bias, such as wind-induced under catch, resulting in underestimation of precipitation. Nonetheless, using a constant correction factor may be a simplification, and the decrease in KGE during validation may be indicative of model overfitting or unrealistic input estimates, which may not be suitable for other time periods. Furthermore, it is plausible that the precipitation correction factors may be time-dependent, and hence stable NWP results may be more reliable. Incorporating different input datasets resulted in a wide range of simulated discharge, particularly in small-scale catchments such as Usma and Grunnåi, corroborating the notion that precipitation data has a substantial influence on the uncertainty of hydrological models, as suggested by Bárdossy & Anwar, 2022.

Evaluating the water balances of the model results enabled assessment of HYPE's capability in handling various input datasets. The findings indicated that HYPE compensates for precipitation underestimation by estimating low evapotranspiration levels to prevent water losses from the catchment. This pattern was observed across all catchments when seNorge2 was used as input data. This phenomenon may be attributed to the absence of actual evapotranspiration when no water is available in the system. However, considering that the catchments are known to have marsh areas, they should not be dry during summer.

Using the corrected version of seNorge2 led to an improvement in the water balance by increasing the amount of water in the catchment. As a result, the model results were closer to the observed values in terms of volume, and even the evapotranspiration estimates aligned more closely with MOD16 values. Seasonal analysis demonstrated that seNorge2 consistently underestimated discharge, particularly during winter.

3.4 Water temperature results

We compare simulated water temperatures in the timeframe 2012 to 2018 at the stations Haga bru (**Figure 13**) and Eggafoss (**Figure 14**).The model fitness is slightly higher at Haga bru for all metrics (**Table 11**). At both stations, simulated temperatures follow the seasonal trend and have a reasonable goodness-of-fit. During the winter and throughout the melting period from May to July, water temperature is underestimated. There is a time lag of temperature increase at Eggafoss, especially pronounced in early snowmelt periods. Modelled water temperature variance is much lower than observed, and peak temperatures are underestimated by a few degrees.

Table 11 *Performance of water temperature model in Gaula catchment*

Station	KGE	NSE	R^2	Mean absolute error	Percent bias
Haga bru	0.89	0.88	0.88	1.30	1.2 %
Eggafoss	0.79	0.86	0.88	1.44	-15.0 %

Figure 13 *Annual water temperature regime – Haga bru 122.28. The lines represent the mean values, and the bands show the variation between 25th and 75th percentile.*

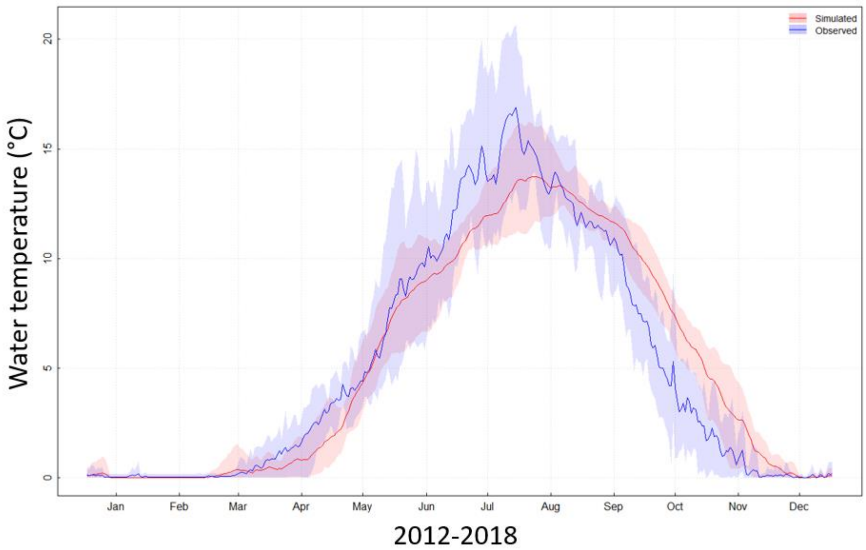
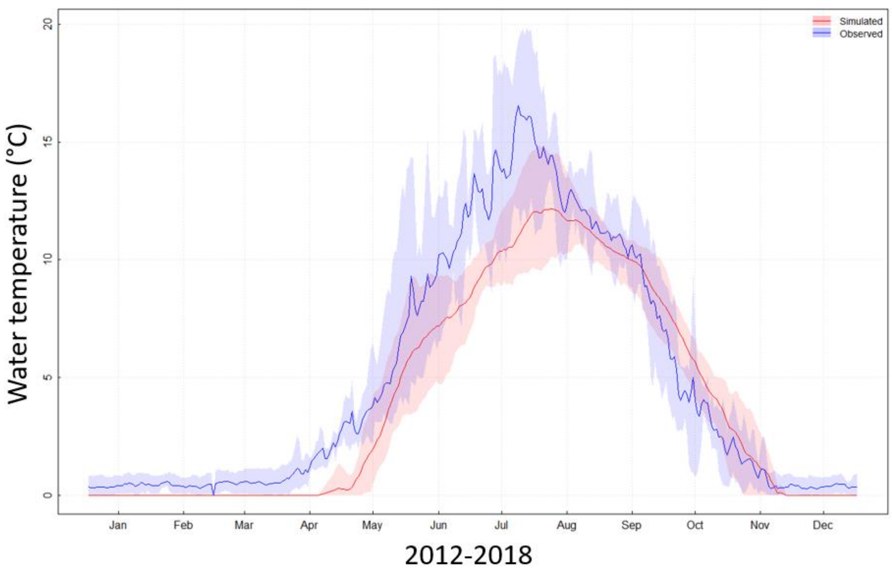


Figure 14*. Annual water temperature regime – Eggafoss 122.11. The lines represented the mean values and the bands show the variation between 25th and 75th percentile.*



4 Discussion

The model delineation is good for medium-sized and large catchments. It is feasible to refine the delineation using WHIST, especially updating the set-up for whole Norway with more refined delineations stemming from individual, detail studies of individual catchments. Our results suggest that the main modules for improving performance in the distributed setup are icatch, stream depth, and lake rating curves, as evidenced by single catchment set-ups both with hourly time step modeling and lake calibration.

We found that KGE for water levels is generally higher than for flow, which is expected given that lakes respond more slowly to changes than rivers, resulting in lower variability in the signal.

We note that at the time of writing this report, validation of all timeframes is incomplete, and a systematic sensitivity analysis has not been conducted. However, a formalized uncertainty analysis is necessary to estimate the model uncertainty.

Regarding our goals, we were able to achieve spatially and temporally relevant output for catchments larger than 100 km^2, and our testing of hourly time steps showed that the model is stable for shorter time steps with recalibration. However, only approximately 20% of the total modeled domain was used for calibration catchments, so our goal of relevant output spatially distributed and of feasible accuracy was only partly achieved.

5 Conclusions and further steps

Our results demonstrate that the model can reliably reproduce regionalized discharge in rivers and water levels in lakes throughout most of the modeled domain, although it is not able to accurately simulate low-flow conditions. While we were able to successfully model water temperature in selected catchments, the model still has flaws, such as early freezing, which may be related to the underestimation of low flows.

The regional calibration (chapter 2.5.3) should be finished for all zones. Testing of larger ensemble sizes and more model runs for automatic calibration.

Further steps can include identifying model modules to improve low flow conditions and overestimation of freezing, for example an estimation of water covered area based on updated GIS data instead of an estimation by sub-catchment area. In addition, set-ups with modules using solar radiation data forcing instead of day-degree-method for evapotranspiration could be tested, and evapotranspiration compared with MODIS16 data.

More regulated catchments can be detailed such as in the Gaula set-up, to allow for detailed assessment of hydrological impact and water temperature impact of hydropower. In a more complex study, implantation of hydropower regulation can be automized using spatial data about the physical infrastructure and operational information such as environmental release flows, ramping and water-level restrictions from concession documents or other data sources.

Further, we can use bias-adjusted climate data to predict the combined impact of climate change and hydropower regulation in catchments.

6 References

Adera, A. G., Alfredsen, N. K., Bakken, N. T. H., Energi, S., & Schönfelder, L. H. (2018). *Assessment of the Hydrological Model HYPE for Southern Norway*.

Allen, R. G. (1998). Crop Evapotranspiration-Guideline for computing crop water requirements. *Irrigation and Drain*, *56*, 300.

Arheimer, B., & Lindström, G. (2013). Implementing the EU water framework directive in Sweden. *Runoff Predictions in Ungauged Basins–Synthesis across Processes, Places and Scales, Edited by: Blöschl, G., Sivapalan, M., Wagener, T., and Viglione, A*, 353–359.

Arheimer, B., Pimentel, R., Isberg, K., Crochemore, L., Andersson, J. C., Hasan, A., & Pineda, L. (2020). Global catchment modelling using World-Wide HYPE (WWH), open data, and stepwise parameter estimation. *Hydrology and Earth System Sciences*, *24*(2), 535–559.

Bárdossy, A., & Anwar, F. (2022). *Why our rainfall-runoff models keep underestimating the peak flows?* [Preprint]. Catchment hydrology/Uncertainty analysis. https://doi.org/10.5194/hess-2022-281

Bartholomé, E., & Belward, A. S. (2005). GLC2000: A new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing*, *26*(9), 1959–1977. https://doi.org/10.1080/01431160412331291297

Bergstroem, S. (1975). The development of a snow routine for the HBV-2 model. *Hydrology Research*, *6*(2), 73.

EEA. (2018). *Corine Land Cover (CLC)*. https://land.copernicus.eu

FAO/IIASA/ISRIC/ISS-CAS/JRC. (n.d.). *Harmonized World Soil Database* (1.1) [Data set]. FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Gottschalk, L., Jensen, J. L., Lundquist, D., Solantie, R., & Tollan, A. (1979). Hydrologic Regions in the Nordic Countries. *Hydrology Research*, *10*(5), 273–286. https://doi.org/10.2166/nh.1979.0010

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, *377*(1–2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Hargreaves, G. H., & Allen, R. G. (2003). History and Evaluation of Hargreaves Evapotranspiration Equation. *Journal of Irrigation and Drainage Engineering*, *129*(1), 53–63. https://doi.org/10.1061/(ASCE)0733-9437(2003)129:1(53)

Hargreaves, G. H., & Samani, Z. A. (1985). Reference Crop Evapotranspiration from Temperature. *Applied Engineering in Agriculture*, *1*(2), 96–99. https://doi.org/10.13031/2013.26773

Huang, S., Eisner, S., Magnusson, J. O., Lussana, C., Yang, X., & Beldring, S. (2019). Improvements of the spatially distributed hydrological modelling using the HBV model at 1 km resolution for Norway. *Journal of Hydrology*, *577*, 123585. https://doi.org/10.1016/j.jhydrol.2019.03.051

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Lindsay, J. B. (2016). Whitebox GAT: A case study in geomorphometric analysis. *Computers & Geosciences*, *95*, 75–84. https://doi.org/10.1016/j.cageo.2016.07.003

Lindström, G. (2016). Lake water levels for calibration of the S-HYPE model. *Hydrology Research*, *47*(4), 672–682. https://doi.org/10.2166/nh.2016.019

Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, *201*(1–4), 272–288. https://doi.org/10.1016/S0022-1694(97)00041-3

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*(1), 235–249. https://doi.org/10.5194/essd-10-235-2018

Pierce, D. (2023). *ncdf4: Interface to Unidata netCDF Format Data Files*. https://CRAN.R-project.org/package=ncdf4

R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Revisjonsdokument for reguleringer i Lundesokna og Sokna. (2019). https://www.nve.no/konsesjon/konsesjonssaker/

Saldana Espinoza, C. I. (2022). *Study of unregulated flow conditions in Norwegian rivers- Strategy for improving lake outflow using HYPE model* [Other]. University of Stuttgart.

Santos, L., Thirel, G., & Perrin, C. (2018). Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrology and Earth System Sciences*, *22*(8), 4583–4591. https://doi.org/10.5194/hess-22-4583-2018

Schönfelder, L. (2017). *Performance assessment of the semi-distributed hydrological model HYPE for central Norway* [NTNU]. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2454731

Sienel, J. (2022). *Evaluation of precipitation products in Norway*. University of Stuttgart.

SMHI. (2022). *The HYPE wiki pages*. SMHI. http://www.smhi.net/hype/wiki/doku.php

Strömbäck, L., Arheimer, B., Lindström, G., Donnelly, C., Gustafsson, D., Isberg, K., Pechlivanidis, I., Andersson, J., & Pers, C. (2013). *The Importance of Open Data and Software for Large Scale Hydrological Modelling*.

Zarei, A. R., Zare, S., & Parsamehr, A. H. (2015). Comparison of several methods to estimate reference evapotranspiration. *West African Journal of Applied Ecology*, *23*(2), 17–25.

Appendix

Table 12 Parameter groups used in automatic calibration

Group name	Parameters
Soil	wcfc1, wcfc2, wcfc3,wcwp1, wcwp2,wcwp3, wcep1, wcep2, wcep3, mperc1, mperc2, sfrost, rrcs1, rrcs2, rrcs3, srrate, macrate, mactrinf, mactrsm, frost, srrcs, surfmem, ttrig, treda, tredb, depthrel
Snow	ttpd, ttpi, deepmem, cmlt, ttmp, fscmax, fsck1, fsceff, fscdistmax, fscdist0, fscdist1, sdnsnew, snowdensdt, fsclim, fsckexp, pcusnow
Lakes	gratk, gratp, grata
Evapotranspiration and other	lp, epotdist, cevp, rivvel, damp, deadl, deadm, pcurain
Individual lakes LakeData.txt	rate, exp

Table 13. Parameters groups calibrated manually / left uncalibrated.

Parameter type	File location	Parameter names (# of individual parameters)
Soil depths	GeoClass.txt	Depth1, Depth2 (10)
Stream depths	GeoClass.txt	Stream-depth (5)
Fraction of runoff which passes through local lake	GeoData.txt	icatch (16581)
Lake rating curve	LakeData.txt	Rate, exp, w0 (3 per *olake*)

Table 14 Calibration catchments with calibration performance for the single-step calibration approach

stID	Station name	Area [km2]	KGE	NSE	Bias [%]
212.49.0	Halsnes	145	0,41	0,44	-1,56
133.7.0	Krinsvatn (Kringsvatnet)	206	0,43	0,55	-3,92
122.17.0	Hugdal bru	546	0,8	0,71	-1,46
109.9.0	Driva v/Risefoss	746	0,72	0,69	-2,74
212.27.0	Eibyelva v/Hammeren	625	0,7	0,64	-1,56
2.479.0	Li bru	157	0,59	0,43	-1,15
75.28.0	Feigumfoss	48	0,52	0,68	-0,91
72.5.0	Brekke bru	268	0,46	0,6	-7,48
212.27.0	Eibyelva v/Hammeren	625	0,7	0,64	-1,56
2.11.0	Narsjø	119	0,8	0,63	-0,21
161.7.0	Tollåga	225	0,56	0,59	-2,73
311.460.0	Engeren	395	0,52	0,11	-1,76
62.5.0	Bulken (Vangsvatnet)	1092	0,7	0,79	-15,49
247.3.0	Karpelva	129	0,69	0,7	-0,52
22.22.0	Søgne	204	0,82	0,73	0,26
311.6.0	Nybergsund	4421	0,35	0,04	-15,66
75.23.0	Krokenelv	46	0,49	0,64	-0,78

2.279.0	Kråkfoss	435	0,73	0,7	-1,74
62.10.0	Myrkdalsvatn	158	0,63	0,67	-3,08
122.11.0	Eggafoss	655	0,58	0,67	-5,1
308.1.0	Lenglingen	453	0,38	0,51	-7,04
139.17.0	Bertnem	5160	0,47	0,36	-113,54
28.7.0	Haugland	139	0,85	0,82	-0,85
223.2.0	Lombola	877	0,45	0,45	-6,1
12.97.0	Bergheim	4233	0,23	-0,64	-28,7
50.13.0	Bjoreio	263	0,35	0,38	-4,35
16.128.0	Austbygdåi	344	0,84	0,84	-0,92
2.290.0	Brustuen	254	0,27	0,4	-5,78
162.3.0	Skarsvatn	145	0,64	0,63	-1,43
139.26.0	Embrethølen	494	0,3	0,47	-11,93
127.11.0	Veravatn	175	0,47	0,61	-3,05
155.27.0	Lendingosen (Varnvatnet)	157	0,53	0,65	-2,37
2.614.0	Rosten	1833	0,59	0,59	-9,37
109.42.0	Driva v/Elverhøy bru	2438	0,58	0,49	-26,94
12.70.0	Etna	570	0,72	0,68	1,56
206.3.0	Manndalen bru	200	0,42	0,43	-2,47
12.192.0	Sundbyfoss	75	0,79	0,69	-0,25
74.1.0	Årdalsvatn	979	0,42	0,23	-25,23
22.4.0	Kjølemo	1758	0,64	0,51	-13,64
83.6.0	Byttevatn	105	0,49	0,58	-3,69
2.291.0	Tora	262	0,2	0,39	-8,01
97.1.0	Fetvatn (Fitjavatnet)	89	0,35	0,31	-3,6
151.15.0	Nervoll	655	0,6	0,69	-9,57
2.284.0	Sælatunga	458	0,3	0,39	-4,82
12.209.0	Urula	554	0,72	0,73	-2,99
3.22.0	Høgfoss	299	0,6	0,42	0,24
62.18.0	Svartavatn	72	0,5	0,54	-1,99
313.10.0	Magnor	358	0,48	0,18	-0,91
123.31.0	Kjeldstad i Garbergelva	143	0,19	0,31	-4,17
124.2.0	Høggås bru	494	0,25	0,23	-11,19
12.215.0	Storeskar	119	0,65	0,78	-0,99
100.1.0	Valldøla v/Alstad	226	0,43	0,52	-5,83
12.178.0	Eggedal	311	0,83	0,83	-0,96
24.8.0	Møska (Skolandsvatnet)	121	0,78	0,64	0,19
139.25.0	Skjellbreivatn	546	0,53	0,59	-6,62
24.9.0	Tingvatn (Lygne)	272	0,73	0,72	-3,57
2.32.0	Atnasjø	463	0,64	0,48	-2,55
12.207.0	Vinde-elv	269	0,83	0,7	0,41
2.634.0	Lena	184	0,82	0,72	-0,3
15.53.0	Borgåi	94	0,75	0,59	0,17

212.49.0	Halsnes	145	0,41	0,44	-1,56
2.28.0	Aulestad	870	0,72	0,59	-0,5
20.2.0	Austenå	276	0,57	0,65	-3,99
128.5.0	Støafoss	477	0,32	0,52	-10,17
98.4.0	Øye ndf.	139	0,44	0,45	-4,11
16.127.0	Viertjern	47	0,73	0,73	-0,17
2.129.0	Dølplass	2014	0,4	0,02	5,41
138.1.0	Øyungen	239	0,49	0,65	-3,94
212.1.0	Halsnes	144	0,41	0,44	-1,56
73.21.0	Frostdalen	26	0,29	0,52	-0,53
2.439.0	Kvarstadseter	375	0,74	0,57	0,34
82.4.0	Nautsundvatn	219	0,84	0,86	-2,2
122.14.0	Lillebudal bru	168	0,46	0,57	-2,02
2.102.0	Gjende	376	0,47	0,6	-7,9
2.145.0	Losna	11206	0,65	0,59	-80,01
15.79.0	Orsjoren	1178	0,74	0,71	-3,91
208.3.0	Svartfossberget	1932	0,75	0,68	-2,93
105.1.0	Osenelv v/Øren	138	0,56	0,44	-1,97
2.142.0	Knappom	1643	0,68	0,61	-4,32
2.265.0	Unsetåa	620	0,8	0,6	-0,11
83.2.0	Viksvatn (Hestadfjorden)	508	0,59	0,43	-14,97
36.9.0	Middal	46	0,51	0,63	-1,2
50.1.0	Hølen	231	0,82	0,74	-1,32
117.4.0	Valen (Laksvatnet)	40	0,57	0,6	-0,56
139.35.0	Trangen	852	0,45	0,56	-15,79
122.9.0	Gaulfoss	3086	0,58	0,65	-22,39
234.18.0	Polmak nye	14170	0,73	0,59	4,07
25.24.0	Gjuvvatn	97	0,39	0,45	-3,38
234.13.0	Vækkava, Iesjokka	2079	0,63	0,54	-8,6
212.1.0	Halsnes	144	0,41	0,44	-1,56
41.1.0	Stordalsvatn	131	0,48	0,49	-6,21
12.114.0	Garhammerfoss	492	0,7	0,72	-2,97
2.268.0	Akslen	789	0,4	0,5	-11,75
196.35.0	Malangsfoss	3111	0,49	0,48	-37,55
152.4.0	Fustvatn	526	0,52	0,52	-13,01
74.18.0	Fornabu	53	0,55	0,69	-0,64
2.604.0	Elverum	15450	0,44	0,39	-38,04

www.hydrocen.no

ISSN: 2535-5392
ISBN: 978-82-93602-32-3

HydroCen
v/ Vannkraftlaboratoriet, NTNU
Alfred Getz vei 4,
Gløshaugen, Trondheim

www.hydrocen.no
HydroCen
@FMEHydroCen