Data Descriptor

Title

Hydrologic outputs generated with the 2023 calibrated version of GEM-Hydro over the Great-Lakes.

Authors

Étienne Gaborit[1], Juliane Mai[2,3,4], Daniel Princz[5], Hongren Shen[6], Vincent Vionnet[1], Bryan Tolson[6], and Vincent Fortin[1]

Affiliations

1. Meteorological Research Division, Environment and Climate Change Canada, QC, Canada
2. While at: Computational Hydrosystems, Helmholtz Centre for Environmental Research – Leipzig, Germany
3. While at: Center for Scalable Data Analytics and Artificial Intelligence – Leipzig, Germany
4. Now at: Department of Earth and Environmental Science, University of Waterloo, ON, Canada
5. National Hydrologic Services, Environment and Climate Change Canada, SA, Canada
6. Department of Civil and Environmental Engineering, University of Waterloo, ON, Canada

corresponding author(s): Étienne Gaborit (Etienne.Gaborit@ec.gc.ca)

Abstract

The dataset described in this document contains outputs from a calibrated version of the GEM-Hydro hydrologic model developed at Environment and Climate Change Canada (ECCC), over the Great Lakes and Ottawa river basins, and for the period 2001-2018 included. This dataset is available on the Federated Research Data Repository.
The outputs consist of all variables (hourly fluxes and state variables) related to the water balance of the land-surface scheme of GEM-Hydro, as well as mean daily streamflow time-series (observed and simulated) at 212 gauge stations within the region of interest.
To produce these outputs, a calibrated version of the GEM-Hydro model was run in open-loop mode (no land-surface or streamflow assimilation performed), driven with atmospheric forcings coming from the ECCC Canadian Surface Reanalysis (CaSR version 2.1).
GEM-Hydro is shown here to be able to perform satisfactory simulations of various hydrologic variables, when compared to reference datasets. The hydrologic variables of this dataset include precipitation, surface runoff, sub-surface runoff (soil lateral flow), soil base drainage, evapo-transpiration, snow water equivalent, soil moisture content for 6 soil layers down to 2m, water stored in the vegetation canopy, and streamflow.
These variables can be used for example to run and calibrate any routing model, compute climatologies, various statistics, or trends for different hydrologic variables over the region of interest, assess the variability of these variables as a function of the local geo-morphology, etc.

Background & Summary

The Great-Lakes basin is a transboundary watershed between the U.S.A. and Canada. It is the largest surface freshwater system on Earth and is home to 37 million people. Together with the Ottawa River basin, they represent a total drainage area of about 965 000 km^2. This system is crucial in terms of freshwater resources. For water resource management and in the context of climate change, it is important to understand the main processes governing the different water balance components in the region, as well as their spatial and temporal variability. Despite observations are being available for several components of the water balance, most of these observations only consist of in-situ measurements, which can be sparse both in space and time while being rarely co-located between various variables observed. Physically-based, distributed hydrologic models can, however, produce consistent, seamless, and continuous (both in space and time) simulations of the different terms of the water balance equation, with a relatively detailed spatial and temporal resolution. These models can also be applied for various applications, such as scenario testing or climate-change impact assessment.

GEM-Hydro (Gaborit et al. 2017[1], Vionnet et al. 2020[2]) is a physically-based, distributed hydrologic model developed at Environment and Climate Change Canada (ECCC). This model is used inside the National Surface and River Prediction System (NSRPS, see Durnford et al. 2021[3]: poster available at https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/383559). The NSRPS is used at ECCC to perform real-time surface and hydrologic analyses and forecasts, both deterministic and probabilistic, over Canada and the main US/Canada transboundary watersheds.

So far, the GEM-Hydro version used in NSRPS is the "default" version of the model, in the sense that none of its internal parameters were calibrated to maximize streamflow or other variables' performances. Despite the default version of GEM-Hydro can have very satisfactory performances in some areas, it can also show very limited performances in others. Agricultural areas are a good example of regions where the default GEM-Hydro version has limited performances with regard to streamflow simulations, because tile drains that are generally installed in such areas to drain any excess of water, are not explicitly represented in the model. Despite the model being physically-based, it cannot accurately represent all the physical processes unfolding in reality (Baroni et al. 2019[4], Budhathoki et al. 2020[5]). Moreover, there are not enough field observations to accurately parameterize the model in all areas of a given domain (Hirpa et al. 2018[6]). Finally, even physically-based models still rely on empirical or conceptual relationships in some parts of the model (Mai 2023[7]), for example to translate soil texture information into soil hydraulic conductivities. For all these reasons, any hydrologic model still needs calibration to achieve its best performances, as illustrated by many recent studies, such as, for example, those of Budhatoki et al. (2020[5]), Hirpa et al. (2018[6]), Bajracharya et al. (2023[8]), Mai 2023[7], Mei et al. 2023[9], and Demirel et al. 2024[10], among many others.

GEM-Hydro has already been calibrated in the past (see Gaborit et al. 2017[1], Mai et al. 2021[11], Mai et al. 2022[12]) for research purposes, mainly as part of the different Great Lakes Runoff Intercomparison Projects (GRIP). During these projects, the tools and methodology used to calibrate GEM-Hydro have continuously evolved. For example, the routing component of GEM-Hydro was replaced during calibration with a simple Unit Hydrograph technique in Gaborit et al. 2017[1] to save computation time, while it was implemented in Mai et al. 2021[11] with a coarse 10-km resolution, before being replaced in Mai et al. 2022[12] by the Raven routing model (still to save computation time). Even after the most recent GRIP project that focused on the whole Great-Lakes region (GRIP-GL, see Mai et al. 2022[12]), improvements were needed to further improve the GEM-Hydro calibration, mainly with regard to surface soil moisture (SSM) in the Lake Erie and Ottawa River basins, and with regard to evapo-transpiration (ET) and snow water equivalent (SWE) in the Ottawa River basin.

The GEM-Hydro outputs shared in the dataset presented here are published on the Federated Research Data Repository (FRDR; Gaborit 2024[13]) and were produced with the most recent calibrated version of GEM-Hydro (referred to as the 2023 calibrated version), at the time of writing. This 2023 calibrated version was obtained by following generally the same calibration methodology as the one used in the GRIP-GL project (see Mai et al. 2022[12] and its supplements for more details), but with several significant differences compared to it. The methodology employed to produce the 2023 calibrated version is explained in detail in the Methods section, but a brief summary of the four main steps involved is presented further down and in Figure 1. However, a brief overview of the GEM-Hydro model is given first, as it may help the reader to better understand the general strategy used here to calibrate GEM-Hydro.

The GEM-Hydro model is composed of two main components. The first one is GEM-Surf (Bernier et al. 2011[14]), which is the surface component of the model. It can represent up to 5 different types of surface covers (or tiles) inside each grid cell, i.e., glaciers, land, water, frozen water, and urban areas. Over land, GEM-Hydro relies on the Soil, Vegetation and Snow (SVS) land-surface scheme (Alavi et al. 2016[15], Husain et al. 2016[16], Leonardini et al. 2020[17], 2021[18]). SVS represents two types of snow-free covers, i.e., high and short vegetation, , and bare ground. In winter, it represents two different snowpacks, i.e., snow over bare ground and short vegetation, and snow under high vegetation. SVS represents a single soil column made of several layers (generally 7, that go down to a 3-m depth). The second GEM-Hydro component is WATROUTE (Kouwen 2010[19]), which is the routing component of the model that is used to convey water in the network of lakes and rivers and to simulate streamflow. WATROUTE is a gridded routing model implemented with a 1-km resolution in GEM-Hydro. It can represent lakes and reservoirs, as well as diversions. To represent regulated reservoirs, WATROUTE relies on the Dynamically Zoned Target Release (DZTR) reservoir model (Yassin et al. 2019[20], Gaborit et al. 2022[21]). Watroute simulates baseflow (the contribution of the aquifers to the surface network) using a conceptual reservoir in each grid-cell, called the Lower Zone Storage (LZS), which is fed with drainage from GEM-Surf and estimates baseflow based on a power function. In GEM-Hydro, GEM-Surf and WATROUTE are one-way coupled, meaning that the routing scheme has no impact on the surface component.

GEM-Hydro is computationally expensive, which is generally the case for physically-based distributed hydrologic models, when implemented over large scales (Baroni et al. 2019[4]). Therefore, when it comes to calibration, where many simulations are required over relatively long time periods, the model's computational time can become a strong limitation for the calibration exercise (see for example Hirpa et al. 2018[6]). In the case of GEM-Hydro, the model is not directly usable for calibration purposes when implemented over large regions, such as the one considered here. The two main reasons for the high computational time of GEM-Hydro are that the surface component runs using successive 24-h integration cycles (a lot of time is spent in input/output processing), and that the routing scheme is gridded, implemented at a 1-km resolution, and not parallelized (all grid points are processed in a sequential manner from the most upstream points to the outlet).

Therefore, the "Modélisation Environnementale communautaire - Surface et Hydrologie" platform (MESH, see Pietroniro et al. 2007[22]) including the SVS land-surface scheme, along with the Raven routing model (Craig et al. 2020[23]) is used for the calibration instead (hereafter referred to as MESH-SVS-Raven). This corresponds to the first step of the general methodology employed here (see Figure 1). MESH-SVS is faster to run compared to the surface component of GEM-Hydro; mainly because it runs over a given time period using a single integration, i.e., not stopping and restarting each day like GEM-Hydro does. Raven routing is much faster than WATROUTE (see Mai et al. 2022[12]) because Raven routing is vector-based, while WATROUTE is a grid-based routing model implemented with a high resolution (1km). The calibrated parameters obtained with MESH-SVS-Raven are then transferred into the actual GEM-Hydro model (step 2 of Figure 1). For these first two steps, the atmospheric forcings, geophysical fields, model configuration, etc. for both the GEM-Hydro and MESH-SVS-Raven models were exactly the same as those used in the GRIP-GL project (Mai et al. 2022[12]). This was done in order to be able to assess the benefit of the changes employed here for the GEM-Hydro calibration methodology, when compared to the one employed during GRIP-GL (see next section). After the second step, a comprehensive evaluation of the calibrated GEM-Hydro version was performed including streamflow performances, but also auxiliary hydrologic variables and near-surface meteorological variables (see "Technical Validation" section). This was done to ensure that the calibration exercise performed here by only maximizing streamflow performances (see “Methods” section) did not degrade other hydrologic variables. If this was the case, changes were then brought to the calibration methodology (see “Methods”) and step 1 was restarted (see Figure 1). After a total of six iterative calibration experiments performed during this study, the resulting GEM-Hydro performances were judged satisfactory, and step 4 was performed (Figure 1). During step 4, the GEM-Hydro setup using the calibrated parameters was modified to use a setup that is close to the one employed in the operational version of NSRPS. This was done to assess the benefit that this calibration exercise could ultimately bring to ECCC’s operational surface and rivers’ analyses and forecasts.

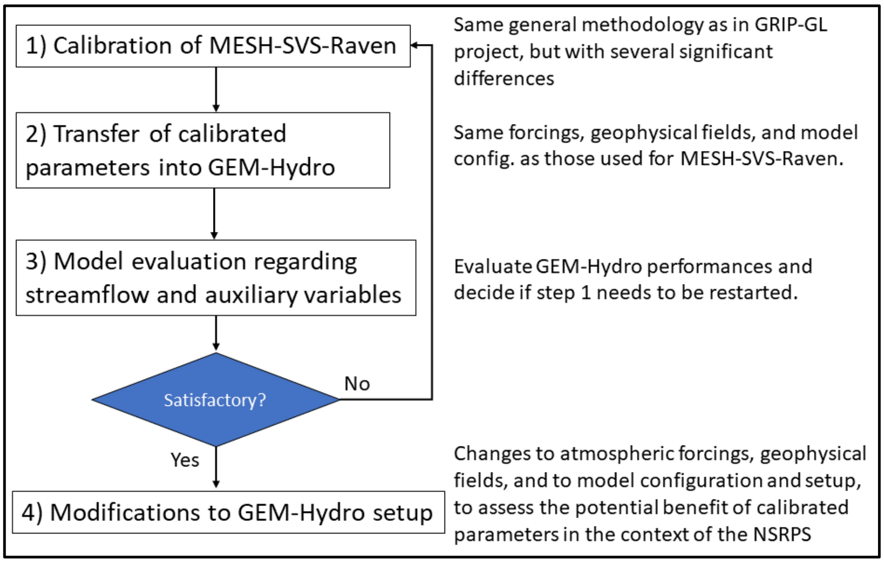

Figure 1: general workflow employed to calibrate some GEM-Hydro model parameters and use them in a setup that is close to the one used in the National Surface and River Prediction System (NSRPS). Note that the setup of step 4 is the one used to produce the outputs that are shared and described in this document.

The GEM-Hydro outputs shared in this dataset were obtained from step 4 of Figure 1, which involves using a setup that is close to the one employed in NSRPS. More details can be found in next section. It is important to note that to produce these outputs, the calibrated version of the GEM-Hydro model was run in open-loop mode (no land-surface or streamflow assimilation performed), driven by atmospheric forcings coming from version 2.1 of the Canadian Surface Reanalysis (CaSR), which relies on the Regional Deterministic Reforecast System (RDRS, see Gasset et al. 2021[24]). Therefore, even if no assimilation was performed for any surface variable or streamflow during these calibrated GEM-Hydro runs, they were forced with atmospheric forcings that come from a reanalysis. This means that, for example, precipitation forcings correspond to precipitation analyses, which consist of short-term forecasted precipitation fields corrected by various sources of precipitation observations. The fact that no assimilation was performed in these GEM-Hydro runs has advantages. For example, the mass of water is conserved in these simulations (from total grid cell precipitation to total streamflow leaving a watershed), whereas the general water balance can be significantly altered in the case of data assimilation, potentially leading to a decrease in model performance regarding streamflow simulations (see for example Garnaud et al. 2021[25]). Another advantage is that without data assimilation, the physical link between the different simulated hydrologic variables is preserved, allowing for example to analyse the characteristics of the relationship between these hydrologic variables.

The outputs shared in this dataset consist of all variables (hourly fluxes and state variables) related to the SVS water balance (land-surface scheme of GEM-Hydro), as well as mean daily streamflow time-series (observed data and simulated data using WATROUTE), at gauge locations across the region of interest. The land surface variables comprise precipitation, surface runoff, sub-surface runoff (soil lateral flow), soil base drainage, evapo-transpiration (ET), snow water equivalent (SWE), soil moisture (SM) content for 6 soil layers down to 2 m, and water stored in the vegetation canopy. The outputs are provided over the Great-Lakes and Ottawa River basins, and over the period from January 2001 to December 2018. The variables can be used, for example, to run and calibrate any routing model, compute climatology, various statistics, or trends for different hydrologic variables over the region of interest, assess the variability of these variables as a function of the local geo-morphology, etc.

As part of the GRIP-GL project (Mai et al. 2022[12]), GEM-Hydro was often the best model regarding SWE and surface SM (SSM) simulations, when compared to other conceptual and lumped models, or to other distributed and physically-based models widely used in North America, but the version presented here is even better regarding these variables. Therefore, it is argued that the hydrologic variables shared in this dataset can be useful to a broad variety of users of the hydrologic community.

Methods

1) Calibration

1.1 – Similarities with the GRIP-GL calibration methodology.

The general GEM-Hydro calibration methodology employed here is the same as in GRIP-GL (see Mai et al. 2022[12]) and includes the main steps described hereafter. MESH-SVS-Raven is used during the calibration exercise in place of GEM-Hydro, to save computation time (see “Background and Summary”). The same geophysical fields and forcings as those used in the GRIP-GL project (Mai et al. 2022[12]) were used with MESH-SVS-Raven during the calibration exercise performed here. The forcings come from the version 2.0 of the Canadian Surface Reanalysis (CaSR), which relies on the Regional Deterministic Reforecast System v2.0 (RDRS, see Gasset et al. 2021[24]). All model runs (i.e., both MESH-SVS-Raven and GEM-Hydro simulations) started on January 1st, 2000, and were initialized with GEM-Hydro outputs corresponding to January 1st of 2014. These GEM-Hydro outputs are the result of an open-loop run of the default version of the model, over multiple years (from 2000 to 2014). The year 2000 was considered a spinup period for the model runs, and was not used for evaluation (outputs are not shared for this spinup year in the dataset described here).

A global calibration approach (Gaborit et al., 2015[26], Demirel et al. 2024[10]) is used for each of the six main subdomains of the region of interest. There is a subdomain for each specific watershed of the five Great Lakes, plus the watershed of the Ottawa River, as shown on Figure 2. This means that the performances at all flow gauges located inside a given subdomain are considered as a whole, at once, for example using the median or the mean of the different stations' performances. The only variable being considered to compute the objective function for a given subdomain is the mean daily streamflow at flow gauges of this subdomain. This strategy of using a global calibration for each of the different subdomains is also referred to as "regional calibration" (Mai et al. 2022[12]), or as “multi-basin calibration” (Demirel et al. 2024[10]). Figure 2 presents the 6 main subdomains and the 212 basins considered in the region.

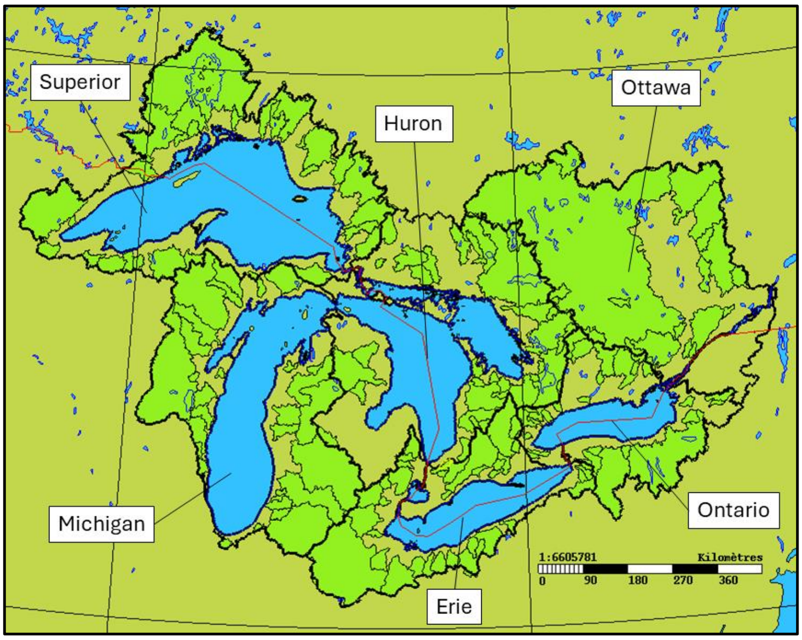

Figure 2: The delineation of the six main subdomains considered here (the 5 Great-Lakes watersheds and the Ottawa River basin), along with the subbasins of the 212 flow gauges used in this study (light green).

The Dynamically Dimensioned Search (DDS) calibration algorithm (Tolson and Shoemaker 2007[27]) is used with a maximum of 240 iterations to perform the calibration of MESH-SVS-Raven. Indeed, DDS is known for its fast convergence (Mai 2023[7]). Moreover, when looking at the evolution of the objective function as a function of the number of iterations during calibration, for any calibration experiment performed in this study, it was clear that the objective function reached an asymptotical behaviour before the maximum number of 240 iterations allowed here (not shown). Note, however, that this number may be insufficient in other calibration contexts.

The calibration parameters mostly consist of multiplying coefficients that are used to multiply the actual model parameters (that generally vary in space) the same way to preserve their original spatial variability. The original actual model parameter values are computed by default by the model (SVS or Raven) based on the geophysical fields provided as input to the models. See Table 1 for information about the multiplying coefficients used as calibration parameters during this work. Moreover, some constraints were applied to the adjusted parameter values actually used in the model, to make sure that the modified values would remain physically coherent. For example, the final albedo values were capped to 1.0, and the final 50% root depth (see table 1) was constrained to remain between a minimal value of 1 cm, and a maximal value of 2cm below the 95% root depth.

Table 1: List of the calibration parameters used, along with a short definition. Init.: Initial value; low,high: lower and upper limits of the interval allowed for a given parameter during calibration. Horiz.: Horizontal; Hydraul.: hydraulic; Cond.: Conductivity; coeff.: coefficient; Agric.: agricultural; LZS: Lower Zone Storage; UH: Unit Hydrograph. See text for the justification of interval limits used for some parameters.

		Init.	low	high	Parameter definition
MESH-SVS parameters	MLTM	1	0.8	1.1	snowmelt rate divider
	GRKMOD	1	1	2.5	horiz. hydraul. Cond. multiplier (all soil layers, non-agric. areas)
	GRKMO_A	1	1	100	horiz. hydraul. Cond. multiplier (layer 5, agric. areas)
	KASMOD	1	1	2.5	vert. hydraul. Cond. multiplier (all soil layers, non-agric. Areas)
	KASMO_A	1	1	5.0	vert. hydraul. Cond. multiplier (first 3 soil layers, agric. areas)
	EVMOD	1	0.8	1.5	evapo-transpiration resistance multiplier

	SUMOD	1	0.7	1.2	sublimation resistance multiplier
	RTMOD	1	1	1.5	95% and 100% root depth mutliplier
	DMOD	1	0.7	1.1	50% root depth multiplier
	AMOD	1	1	1.2	albedo multiplier
	URMO	1	1	1.3	urban area impervious fraction multiplier
Raven params.	FLZCOEFF	2.40E-05	1.00E-07	1.00E-03	LZS multiplicative coeff.
	PWRC	2.8	2	4	LZS power coeff.
	R1NC	1	0.5	2	Mannings' coeff. Multiplier
	GASH	1	0.5	2	Multiplier of the shape of the gamma UH
	GASC	1	0.5	2	Multiplier of the scale of the gamma UH
	LACRWD	1	0.1	1	lake outlet width multiplier

1.2 – Differences with regard to the GRIP-GL calibration methodology.

The differences of the methodology employed here to calibrate GEM-Hydro, as compared to the one used in GRIP-GL, include the following changes. Some bugs related to the reading of some geophysical fields in MESH-SVS were corrected (see Mai et al., 2022[12]). Because of these bugs, the calibrated parameter values obtained during GRIP-GL were not optimal for GEM-Hydro, leading to a significant drop of performances between MESH-SVS-Raven and GEM-Hydro. Since the bugs have been fixed, the performances between the two systems are much closer, as can be seen in the "technical validation" section.

All flow stations used in the GRIP-GL project were used here for calibration, except a few ones. In opposition to the GRIP-GL project, where some stations were not used during calibration in order to perform a spatial validation of the calibrated models, as many stations as possible were used here during calibration. This was done in order for the objective function to be as representative as possible of the performances for all parts of a given subdomain. Moreover, it was not necessary to keep stations for the spatial validation of the model, since the spatial robustness of the calibration methodology employed here has already been demonstrated during the GRIP-GL project (Mai et al. 2022[12]). However, some stations were still discarded during calibration. Indeed, some of the basins considered here have a heavily regulated flow regime, while this regulation was not explicitly represented in the Raven setup used here, resulting for some of these stations in poor model performances with the default version of MESH-SVS-Raven. Therefore, it was necessary to discard these basins during calibration, otherwise they could have misguided the evolution of the calibration algorithm. However, not all basins involving flow regulation were excluded during calibration, because in some cases the impact of the regulation did not prevent the default version of MESH-SVS-Raven from achieving satisfactory simulations. This is the case for example for the station 02KF005 (Ottawa River at Brittania): despite the basin includes about 12 major dams, the regulation still has a limited impact on the total flow of the watershed (especially in spring), due to its large size (90 900 km^2). The list of stations that were excluded from the calibration exercise performed here can be found in Table 2. However, note that all stations were considered when comparing the default and calibrated versions of the GEM-Hydro model (see "Technical validation"). Indeed, many major dams of the Great-Lakes and Ottawa River basins are explicitly represented with the DZTR model, in the WATROUTE routing component of GEM-Hydro. The complete list of the 212 flow stations considered in this study, along with their main basin characteristics, can be found in Mai et al. (2022[12]).

Table 2: list of stations not considered during calibration, along with their (revised) Kling-Gupta Efficiency (KGE, see Kling et al. 2012[22]) performances over the period 2001-2010 for the default version of MESH-SVS-Raven, and the default and calibrated versions of the GEM-Hydro model. The stations highlighted in green are those for which the

regulation is explicitly represented in GEM-Hydro using the DZTR model (Yassin et al. 2019[20]). Note that in the MESH-SVS-Raven setup used here, the DZTR model was not used, for any stations.

Subdomain	Station ID	Watershed size (km^2)	MESH-SVS-Raven	GEM-Hydro default	GEM-Hydro calibrated
Huron	02EB011	4790	-1.58	0.35	0.52
Huron	02DD010	13900	0.47	0.63	0.63
Huron	02DB005	3150	0.11	0.3	0.36
Huron	04136000	2870	-0.44	-1.41	-0.59
Huron	04137500	4500	-0.77	-1.67	-0.81
Ont	02HF002	1280	-0.02	0	-0.2
Ott	02LE025	883	-0.11	0.02	0.08
Sup	02AD012	24700	0.42	0.75	0.7
Sup	02BD002	5310	-0.1	-0.1	-0.35
Sup	02BD007	1950	0.13	0.21	0.01
Sup	02BE002	2880	-0.31	-0.24	-0.48
Sup	04044724	210	0.32	0.05	-0.07
Median of the five "green" stations			0.11	0.35	0.52
Median of other stations			-0.10	-0.10	-0.35

For each subdomain, the objective function considered during calibration consists of the mean of the normalized (revised) Kling-Gupta Efficiency (KGE) criteria (see Kling et al. 2012[28]) across the flow stations of this subdomain. Based on previous calibration experiments performed, it was preferred to use the mean than the median. When using the median, some station performances are actually always not reflected in the objective function, resulting into neglecting their performances. However, when using the mean, all stations' performances are taken into account in the objective function. Therefore, a normalized version of the (revised) KGE criteria was required. It corresponds to the revised KGE, but rescaled such that it falls between 0 and 1. This is done to prevent large negative KGE values obtained for some stations from strongly affecting the mean, which would put too much emphasis on the worst stations. In order to normalize the KGE, Equation 1) below was used:

$$KGEn = \frac{1}{2-KGE} \qquad \text{Equation 1), with}$$

$KGEn =$ normalized KGE criteria (values between 0 and 1)

$KGE =$ revised KGE criteria (Kling et al. 2012[28], values between -∞ and 1).

The calibration was performed over the period 2008-2017 included (2007 used as spin-up), and validation period spans over 2001-2007 included (2000 used as spin-up). Using a more recent calibration period was preferred, given that the calibrated parameters are to be used ultimately in the NSRPS real-time forecasting system, and given that land use / land cover may have changed significantly for some areas of the region, over this period of 18 years in total.

Some calibration parameters used in GRIP-GL were discarded here because they were too sensitive on the resulting auxiliary hydrologic variable performances (see "Technical validation"). The calibration parameters that were included in GRIP-GL but excluded here consist of the multiplying coefficients related to the three SM content thresholds (wilting point, field capacity and saturation), the slope of the soil water retention curve, the soil water suction at saturation, the vegetation leaf-area index, and the vegetation roughness (see Mai et al. 2022[12]). Note that the identification of these problematic parameters was progressively done during the 6 total iterations of this calibration exercise (see point 2 below). The lower and upper limits for some parameter intervals mentioned in Table 1 are also different than the values used in GRIP-GL, because they were also progressively refined during these 6 iterations. The refinement of the parameter intervals that are allowed during calibration was performed either to prevent unrealistic resulting parameter values in the GEM-Hydro model, to limit the degradation sometimes noticed for some flow stations or some auxiliary variables (see Technical validation), or simply to discard some parameter ranges that were never chosen as the best values by the calibration algorithm, with the objective to prevent the calibration algorithm to explore useless regions for some parameter values. The latter for example explains why the lower bound interval limit for some parameters of Table 1 are sometimes equal to the initial value of 1.0.

A new approach to represent the effect of tile drains was employed during calibration: a specific calibration parameter was used to increase the horizontal hydraulic conductivity of the SVS soil layer number 5 under agricultural cover, which is located between the depths of 40 and 100 cm, where agricultural tile drains are generally located. This approach to represent the effect of tile drains in large-scale hydrologic models was suggested by De Schepper et al. (2015[29]). However, in SVS, a unique soil column is used, regardless of the type of vegetation cover. Therefore, a unique multiplying coefficient needs to be used to increase hydraulic conductivity. This is done by computing a weighted average of two multiplying coefficients, one specific to agricultural cover, and one specific to covers other than agricultural (see Table 1), as mentioned in Equation 2 below:

$$GRKM(I,K) = GRKMOD(I) * \left(1.0 - FRAC_{A(I)}\right) + GRKMO\_A(I) * FRAC\_A(I) \qquad \text{Eq. 2), where:}$$

$GRKM(I,K)$: final multiplying coefficient used to adjust the soil horizontal hydraulic conductivity of soil layer K, in grid-cell I. Note that Equation 2 above is only used for K=5 (5th soil layer). For the other layers, $GRKM(I,K) = GRKMOD(I)$

$GRKMOD(I)$: multiplying coefficient used to adjust soil horizontal hydraulic conductivity for vegetation covers other than the agricultural type, in grid-cell I.

$FRAC\_A(I)$: fraction of the land tile occupied by agricultural cover in current grid cell I.

$GRKMO\_A(I)$:multiplying coefficient used to adjust soil horizontal hydraulic conductivity for agricultural vegetation cover, in grid-cell I.

Another new approach to represent the effect of ploughing was employed: a specific calibration parameter was used to increase the vertical hydraulic conductivity under agricultural cover of the SVS soil layers 1 to 3 (i.e., for depths between 0 and 20 cm, that are generally strongly impacted by ploughing). However, for the reasons explained above, this specific coefficient related to agricultural cover needs to be merged with the coefficient related to other covers, following equation 3) below.

$$KASM(I,K) = KASMOD(I) * \left(1.0 - FRAC_{A(I)}\right) + KASM\_A(I) * FRAC\_A(I) \quad \text{Eq. 3), where:}$$

$KASM(I,K)$: final multiplying coefficient used to adjust the soil vertical hydraulic conductivity of soil layer K, in grid-cell I. Note that Equation 3) above is only used for K =1 to 3 (first 3 soil layers). For the other layers, $KASM(I,K) = KASMOD(I)$

$KASMOD(I)$: multiplying coefficient used to adjust the soil vertical hydraulic conductivity for vegetation covers other than the agricultural type, in grid-cell I.

$KASM\_A(I)$: multiplying coefficient used to adjust soil vertical hydraulic conductivity for agricultural vegetation cover, in grid-cell I.

$FRAC\_A(I)$: fraction of the land tile occupied by agricultural cover in current grid cell I.

The final values of the calibrated parameters mentioned in Table 1 are shown for the 6 different subdomains on Figure 3. It can be seen on Figure 3 that some calibrated parameters exhibit some spatial consistency between the domains regarding their evolution, because they were generally all increased (or decreased) compared to their default value, which is the case for example for GRKMOD, KASMOD, RTMOD, DMOD, R1NC, GASH, and LACRWID. See table 1 for a definition of the different calibration parameters. On the opposite, some parameters evolved differently compared to their default value depending on the subdomain, which is the case for MLTM, EVMOD, SUMOD, FLZCOEFF, PWRC, and GASC. In the case of the two parameters related to the LZS (FLZCOEFF and PWRC), their final values for the Lake Huron and Lake Michigan subdomains are very different than those chosen for the other subdomains, which moreover display a very strong coherence between them. This could be the indication of hydrogeologic processes that are specific to the Lake Huron and Lake Michigan subdomains, for example related to the existence of deep (confined) aquifers in these areas. Therefore, the simple LZS conceptual model used in WATROUTE to simulate baseflow may not catch the actual processes occurring for these subdomains in reality, which would probably call for a more complex hydrogeologic model. The fact that the GRKMO_A parameter displays a large spread for calibrated values may be due to the fact that some subdomains contain almost no agricultural areas (like the Lake Huron, Lake Superior and Ottawa River subdomains), making the parameter unsensitive during calibration for these domains. This is also probably the case for the URMO parameter related to the impervious degree of urban areas: only the Lake Ontario and Lake Erie subdomains contain basins that are significantly impacted by urban areas. Finally, some parameters reached the upper or lower limit of their allowed interval (see for example GRKMOD, KASMOD, RTMOD, EVMOD, FLZCOEFF or PWRC), but this is the result of the choices made during the six iterations of the calibration trials performed here, as explained above.

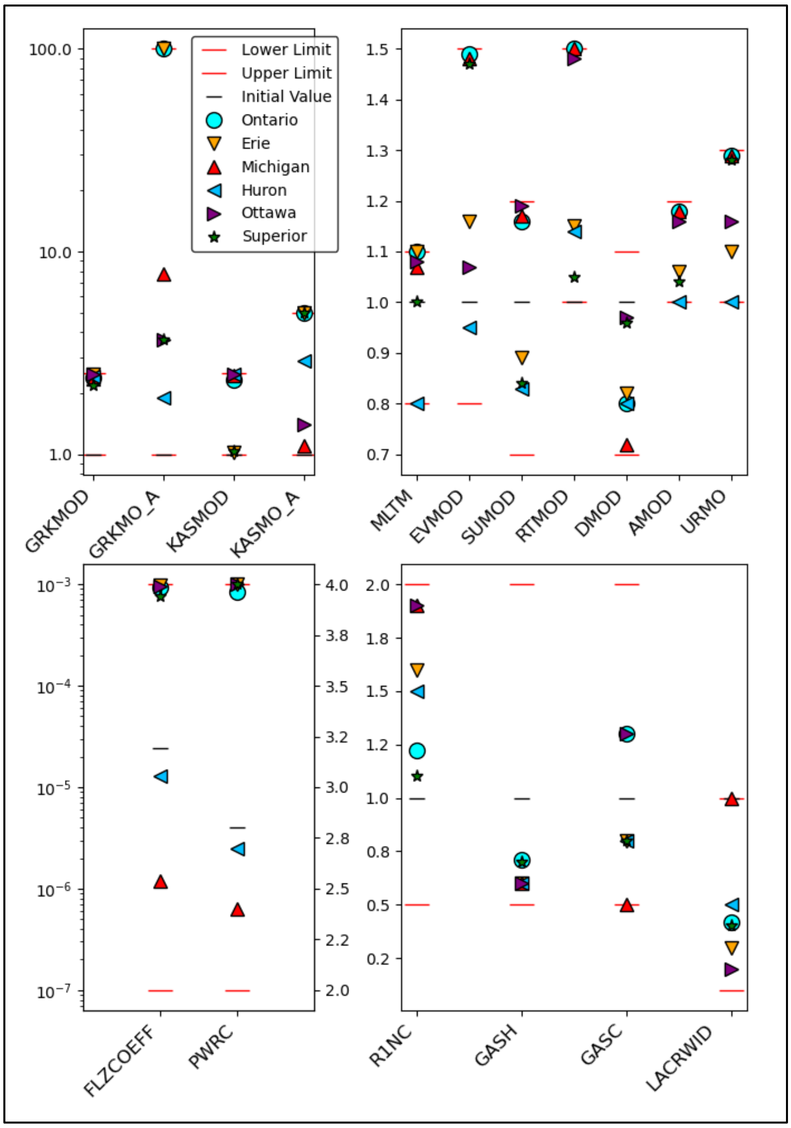

Figure 3: final calibrated parameter values for each subdomain, along with the initial values and the lower and upper interval limits used for each parameter. The top graphs show the SVS parameters, while the bottom graphs show the calibration parameters for the Raven routing model. See Table 1 for a description of the parameters shown here. Note that for the lower left graph, two different y axes are used.

2) transfer of calibrated parameters into GEM-Hydro

Once the calibration was performed with MESH-SVS-Raven, the calibration parameters were transferred into the actual GEM-Hydro model (see step 2 of Figure 1), with no change to the configuration, setup, forcings, etc., such that differences between the calibrated MESH-SVS-Raven and GEM-Hydro models would mainly come from the differences related to the change of the routing model (i.e., from Raven in MESH-SVS-Raven to WATROUTE in GEM-Hydro). This was done in order to ensure that the two different modelling platforms were leading to similar results, and that the calibrated parameters obtained with MESH-SVS-Raven were appropriate for use in GEM-Hydro. See the technical validation section for the differences between MESH-SVS-Raven and GEM-Hydro performances.

3) Model evaluation regarding streamflow and auxiliary variables

At this stage, a comprehensive evaluation of the GEM-Hydro auxiliary hydrologic variables and near-surface variables (SWE, SSM, ET, 2-m temperature and dew point, 10-m wind speed) was performed (see Figure 1 and technical validation) in order to make sure that the calibration exercise, only focused on maximizing streamflow performances, did not degrade hydrologic or near-surface variables, when compared to the performances of the default GEM-Hydro version. If this was the case, then the first step of Figure 1 was restarted after changes were brought to the calibration methodology, for example by removing some calibration parameters that were judged too sensitive on the auxiliary hydrologic or near-surface variables, by refining some intervals allowed for calibration parameter values, or by changing other aspects of the calibration methodology. Note that the calibration methodology described in this document corresponds to the “last” iteration of this calibration exercise, which was obtained after six iterations of the cycle mentioned on Figure 1 for the first three steps of the methodology employed here. Making sure that hydrologic variables other than streamflow, as well as surface variables and fluxes, are not degraded compared to the default version of the model is important for two main reasons. The first reason is related to the fact that the goal of this calibration exercise is to use a calibrated GEM-Hydro version in the NSRPS, where various data are being assimilated to correct several variables, such as snow cover, SSM, and surface temperatures. Therefore, simulation performances need at least to be maintained between the default and calibrated versions of GEM-Hydro for these variables, otherwise the assimilation process may be significantly altered. The second reason has to do with the ultimate goal of the SVS land-surface scheme, which is to be used directly in the atmospheric systems used at ECCC, with the vision of using the same systems both for Numerical Weather Prediction (NWP) and hydrologic forecasts. As such, calibrating SVS to optimize streamflow performances should not lead to degrading the surface temperatures and fluxes, otherwise this would have a negative impact on weather forecasts.

Finally, it is important to emphasize that Raven and WATROUTE are two different routing schemes. Therefore, even if some parameters calibrated using Raven could be directly transferred into WATROUTE (such as the two parameters FLZCOEFF and PWRC related to the groundwater discharge computation, see Table 1), the others could not, because some processes were not represented the same way in the two routing schemes. Therefore, the other WATROUTE parameters remained unchanged compared to the default version of the model. Nevertheless, it was tried to further tune the WATROUTE Manning roughness coefficients, by manually adjusting these values for each of the 6 subdomains with the calibrated GEM-Hydro version. However, since no significant performance gain could be further achieved this way, it was preferred to use the default WATROUTE values for these parameters.

4) Modifications to the GEM-Hydro setup

Then, significant changes were brought to the GEM-Hydro setup using the calibrated parameters (see step 4 of Figure 1), to assess the potential benefit that the calibrated parameters could bring when employed with a GEM-Hydro model configuration that is close to the one employed in the NSRPS system. This modified setup was the one used to produce the GEM-Hydro outputs that are being shared in this dataset. More details about this “final” GEM-Hydro setup are provided below. A comprehensive evaluation of the GEM-Hydro performances obtained after this step was performed to make sure that no significant degradation was noticed for streamflow, auxiliary hydrologic and near-surface variables (see “technical validation”), when compared to the GEM-Hydro setup used in step 2. Table 3 below summarizes the main modifications brought to the GEM-Hydro setup during this step.

Table 3: Main differences between the GEM-Hydro setup used during step 2 of Figure 1 (the “GRIP-GL” setup) and the one used during step 4 (the “NSRPS” setup). [1]: see Mai et al. (2021[12]) for more information about the geophysical

fields used during the GRIP-GL project. Precip. PPM: precipitation phase partitioning method. See text for more details about the NSRPS setup. See Figure 1 for the different steps of the methodology employed here.

	Forcings	Calibrated parameters	Geophysical fields	Modelling option	Model version
GRIP-GL setup (step 2)	CaSR v2.0	Fixed for each subdomain	GRIP-GL	Precip. PPM: 0-degree threshold	GEM-Surf 6.1.2
NSRPS setup (step 4)	CaSR v2.1	Smoothed at subdomain boundaries	NSRPS	Precip. PPM: Harder and Pomeroy (2013[30])	GEM-Surf 6.2.0

4.1- Changes to the surface component

GEM-Surf version 6.2.0 was used. The different GEM-Surf options of this "final" GEM-Hydro setup are not described here in detail. We refer to the readme file of the dataset (Gaborit 2024[13]) for the list of the options used in GEM-Surf. One important change, however, is that this final setup is using the precipitation phase partitioning method (PPM) of Harder and Pomeroy (2013[30]) whereas the previous setup used a 0°C. threshold on air temperature to split between rainfall and snowfall. Using a humidity-based PPM such as Harder and Pomeroy (2013[30]) strongly improved the ability of GEM-Hydro to predict the precipitation phase (Vionnet et al. 2022[31]). Atmospheric forcings come from version 2.1 of the Canadian Surface Reanalysis (CaSR) (Gasset et al. 2021[24]). Regarding the geophysical fields, the final GEM-Surf setup relies on the following data sources, which for some of them differ from the sources used during GRIP-GL (see Mai et al. 2022[12]).

- The Global Multi-resolution Terrain Elevation Data version 2010 (GMTED 2010, see USGS, 2010[32]) was used for surface topography (elevation, slope) and WATROUTE elevation data (different from the one used in GRIP-GL).
- The Climate-Change Initiative – Land Cover dataset version 2015 (CCI-LC 2015, see ESA, 2015[33]) was used for land use / land cover in GEM-Surf and WATROUTE (different from the one used in GRIP-GL).
- The Global Soil Dataset for Earth system modelling (GSDE, see Shangguan et al. 2014[34]) was used for soil texture.
- The National Hydrographic Network (NHN, see Natural Resources Canada, 2020[35]) and the National Hydrographic Dataset (NHD, see USGS, 2021[36]) were used for drainage density.
- HydroSHEDS 30 arcsec. (~1km) resolution (HydroSHEDS, 2021[37]) was used for WATROUTE flow direction grids.

However, in opposition to the actual NSRPS configuration, and similarly to the GEM-Hydro version used during step 2 of Figure 1 (the "GRIP-GL" configuration), subgrid-scale lakes were still deactivated during step 4. This means that when a grid cell contains less than 100% water, the water tile inside that grid-cell is not considered by the model but is replaced by the other surface tiles of the grid cell while preserving their relative importance. Grid-cells containing 100% of water were not modified as no other surface tile could replace the water tile. This "filtering" of water surfaces was also not done for grid cells located around pixels containing 100% water. This filtering is needed because with the model currently used in GEM-Hydro to represent water surfaces, an external source has to be used for water temperature and ice fraction. Generally, an ECCC internal analysis is used for that, but it is not available over the whole period of this study. Therefore, ERA-Interim was used to provide this information, for consistency throughout the whole period. However, based on tests performed with this source, it is not judged reliable in terms of the resulting evaporation simulated over water with GEM-Hydro. Therefore, it was preferred to neglect the water portion of the grid-cells where possible, leading to filtering out the subgrid-scale lakes from the GEM-Hydro setups used during this study. This is why only the hydrologic variables over land are provided in this dataset, and not over water surfaces (see data records). Note that neglecting the subgrid-scale lakes in GEM-Hydro has a limited impact on the resulting streamflow simulations of the region, based on tests previously performed (not shown here).

The GEM-Surf calibrated simulations were performed using a single model setup covering all of the geographic region of interest (i.e., all of the 6 subdomains were included in this single setup). To do so, the calibration parameters (i.e. the multiplying coefficients of Table 1) were provided as 2-D input fields to GEM-Surf (similarly to other static geophysical fields). However, to avoid abrupt parameter changes at subdomain boundaries, the 2-D parameter fields were smoothed. Not doing so could ultimately create abrupt changes in some surface fluxes or variables at these boundaries, which is not desirable in the (future) context of coupling this calibrated version with an atmospheric model. To smooth the calibrated parameter values, the 2-D fields of the fixed parameters values for each subdomain were combined and then aggregated by a factor of 3 (i.e., decreasing the resolution by a factor of 3), before being bilinearly interpolated back on the original grid resolution. Figure 4 shows an example smoothed 2-D field for the GRKMOD calibrated parameter.

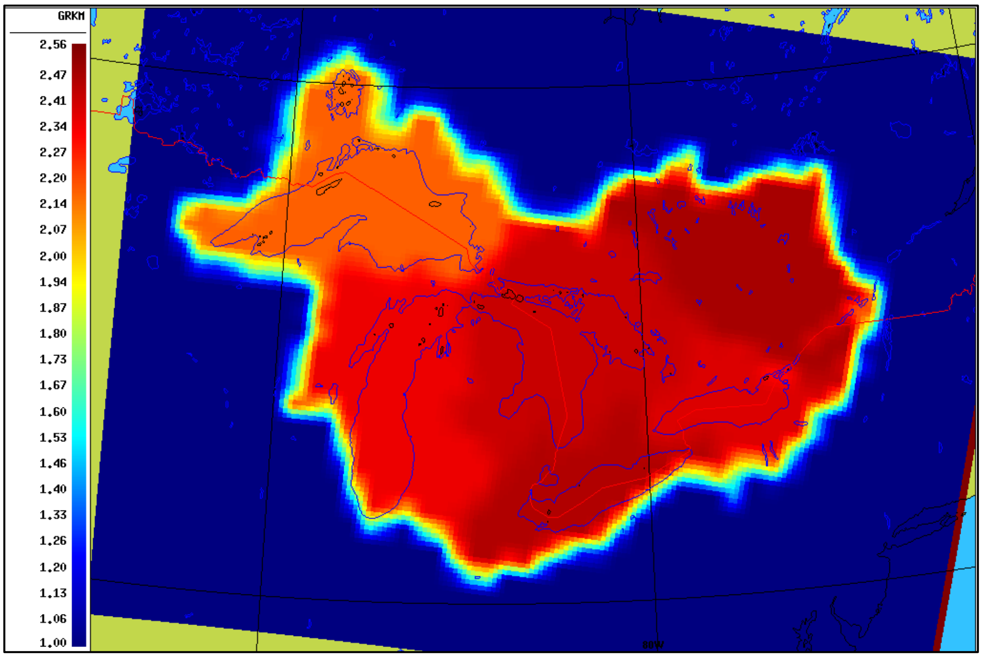

Figure 4: example smoothed 2-D field of the GRKMOD calibrated parameter values that were provided as input to GEM-Hydro when using a single model setup over the full Great-Lakes and Ottawa River domains. Outside of these domains, the default (uncalibrated) value of 1.0 is applied to this parameter. See Table 1 for the definition of the GRKMOD parameter.

4.2- changes to WATROUTE:

The ECCC version 3.4 of WATROUTE was used. Similarly to the GEM-Surf component, as part of the “final” GEM-Hydro open-loop run, WATROUTE was run using a single setup over the full Great Lakes and Ottawa River region. To do so, the two parameters that were calibrated with Raven and transferred into WATROUTE were also provided to WATROUTE as 2-D static fields. Note that in this case however, no smoothing of the parameter values was performed at subdomain boundaries, because in the current GEM-Hydro implementation, the surface and routing components of GEM-Hydro are one-way coupled, such that the routing component cannot have any impact on the surface component (and therefore on the atmospheric model).

5) Methods for processing the GEM-Hydro outputs:

5.1- Surface variables

- Fluxes: PR, ACWF, TRAF, ALAT, and O1 (in mm/h, see Data records):

GEM-Hydro runs over long periods by performing 24-h cycles of continuous integrations, between 12:00 UTC and 12:00 UTC the day after. For each of these 24-h cycles, the raw output fluxes consist of accumulated values in mm since the start of the 24-h integration. These raw outputs were then processed in order to compute the "decumulated" fluxes, such that each flux provided in this dataset consists of the quantity of water (in units of kg/m^2, or mm) over the hour preceding the date mentioned. All dates of the GEM-Hydro surface variables shared in this dataset correspond to time in the Universal Time Coordinated (UTC) format.

- Snow Water Equivalent (SWE, in mm):

GEM-Hydro does not directly output the mean SWE over the land part of a grid-cell. GEM-Hydro (SVS) does simulate two different snowpacks over the land area of a grid cell: one under high vegetation, and one over bare ground/ short vegetation. The mean land SWE in a grid cell was computed from the raw outputs using Equation 4 below:

$$\mathrm{SWE} = (\mathrm{SNDP} * \mathrm{SNDN} * 0.01 + \mathrm{WSN}) * (1. - \mathrm{VEGH}) + (\mathrm{SVDP} * \mathrm{SVDN} * 0.01 + \mathrm{WSV}) * \mathrm{VEGH}$$ Eq. 4, where:

SNDP, SVDP: snowpack depth (cm) respectively over bare ground + short vegetation, and under high vegetation.

SNDN, SVDN: snowpack density (kg/m^3) respectively over bare ground + short vegetation, and under high vegetation.

WSN, WSV: liquid water stored in the snowpack (mm), respectively over bare ground + short vegetation, and under high vegetation.

VEGH: fraction of the land area of the grid cell that is covered with high vegetation.

- Water stored in/on vegetation (WVEG, in mm):

In SVS, the quantity of water stored in/on vegetation is valid for the fraction of the land surface for which vegetation is above the snowpack. Therefore, the raw WVEG SVS output was modified according to Equation 5 below, such that it corresponds to a height of water (in mm) valid over the whole land surface area, in order to be used directly in Equations 6 and 7 to compute the SVS water balance over the land-surface area of a grid-cell (see Data Records).

$$WVEG = WVEG_{raw} * (1 - PSGL)$$ Equation 5, where:

$WVEG$: final WVEG variable shared in this dataset, and valid over the whole land surface area of a grid-cell (in mm)

$WVEG_{raw}$: raw SVS output corresponding to the quantity of water (in mm) stored in the vegetation fraction of a grid-cell's land tile that is above the snowpack.

$PSGL$: fraction of the land surface that is covered with snow over bare ground and snow over low vegetation.

- Soil moisture (WSL1-6, m^3/m^3):

No manipulation of the raw soil moisture outputs was performed but information is provided here on how to convert the values provided in this dataset. Soil moisture has units of m^3/m^3: it represents the fraction of a given soil layer that is filled with liquid water (the version of SVS used here does not represent freeze/thaw processes). In order to convert soil moisture from m^3/m^3 into soil moisture with units of mm (needed to compute the SVS water balance for example, see Equations 6 and 7), one needs to multiply the soil moisture (m^3/m^3) for a given soil layer by the thickness of this soil layer in mm. Then, one could sum up the soil moisture content over the six soil layers in mm to obtain the total amount of water stored in the SVS soil column, in units of mm. The depth of the SVS soil layers are mentioned in Table 4, but their thickness is mentioned below in mm. The total soil thickness in this GEM-Hydro configuration is equal to 2000mm (2m).

- soil layer 1: depth between 0 and 5 cm, thickness of 50mm
- soil layer 2: depth between 5 and 10 cm, thickness of 50mm
- soil layer 3: depth between 10 and 20 cm, thickness of 100mm
- soil layer 4: depth between 20 and 40 cm, thickness of 200mm
- soil layer 5: for depth between 40 and 100 cm, thickness of 600mm
- soil layer 6: for depth between 100 and 200 cm, thickness of 1000mm

For all of the GEM-Surf outputs (2D gridded fields, so all except streamflow) provided in this dataset (except the WT variable), the fields were then filtered in order to remove the values located outside of the region of interest, i.e. outside of each of the five Great-Lakes watersheds, and outside of the Ottawa River watershed. This was done because the calibrated parameter values only apply inside of these watersheds. Therefore, outside of them, the GEM-Hydro outputs correspond to the default version of the model. It was preferred not to mix outputs corresponding to the default version of the model with outputs corresponding to the calibrated version of the model in the same files included in this dataset. The values filtered out were replaced by "NaN" values. However, the WT variable is not simulated by the model and corresponds to a static field that represents the fraction of the grid-cell occupied by land in the setup used here. It is reminded that subgrid-scale lakes were deactivated here (see point 4.1) such that most of the grid cells outside of the big lakes have a WT value (land fraction) equal to 1.0 here (100% land). Therefore, this field was not filtered out outside of the region of interest.

For all of the GEM-Surf outputs (2D gridded fields, so all except streamflow), the standard files (a binary format used internally at ECCC) were all converted into the netcdf format. All dates associated with the GEM-Surf variables shared in this dataset correspond to time in the Universal Time Coordinated (UTC) format.

5.2- Streamflow:

The WATROUTE component of GEM-Hydro produces streamflow at a 1-km resolution, with an hourly time-step. However, in this dataset, we only provided simulated and observed streamflow time-series for the grid cells containing a flow gauge. Moreover, the hourly flows were converted into mean daily flows because a daily time-step was used to evaluate model performances regarding streamflow. The daily flows shared in this dataset for a given date are valid between 00:00 and 00:00 the day after in local time, i.e. from midnight to midnight, with time always corresponding to the Eastern Daylight Time (EDT, corresponding to UTC minus 4 hours). More information on the 212 flow gauges for which mean daily flows are reported in this dataset can be found in Mai et al (2022[12]), or on the USGS and Water Survey of Canada (WSC) websites mentioned below. The United States' (US) streamflow observations included in this dataset come from the US Geological Survey (USGS) website mentioned below.

https://waterdata.usgs.gov/nwis/dv/?referred_module=sw (accessed on December 1st , 2023)

https://wateroffice.ec.gc.ca/search/historical_e.html (accessed on December 1st, 2023)

The Canadian daily streamflow observations were obtained from the HYDAT database available for download here:
National Water Data Archive: HYDAT - Canada.ca (accessed on December 1st, 2023).

Data Records

The GEM-Hydro model hydrologic outputs shared in this dataset (Gaborit 2024[13]) were published on the Federated Research Data Repository (FRDR) and cover the period from 2001-01-01 to 2018-12-31. The outputs are available over the Canadian/USA watersheds of the Great Lakes, and over the Ottawa River watershed. In other words, they are available over the watersheds of Lake Superior, Lake Michigan, Lake Huron and Georgian Bay watershed, Lake Erie and Lake Saint-Clair watershed, Lake Ontario watershed, and the Ottawa River watershed (Canada, provinces of Ontario/Québec). These outputs consist of all variables (hourly fluxes and state variables) related to the water balance (see Equations 6 and 7 below) of the land-surface tile (SVS) of the surface component of GEM-Hydro (GEM-Surf), and of the mean daily streamflow time-series (observed and simulated with the WATROUTE routing component of GEM-Hydro), at the 212 gauge locations of the region of interest.
The equations of GEM-Hydro land-surface (SVS) water balance can be written as follows for any grid-cell and over any temporal period:

$\Delta S = PR - (ACWF + TRAF + ALAT + O1)$ Equation 6), with:

ΔS: Change in storage (final storage – initial storage) between the final and initial dates of the period being considered, in mm (or kg/m^2).

$PR, ACWF, TRAF, ALAT, O1$: Accumulated values of the different SVS water fluxes (see Table 4) over the period being considered, in mm or kg/m^2.

The equation to compute the SVS water storage for any given date is given below:

$S = WVEG + SWE + \sum_{i=1,6}(WSL_i * H_i)$ Equation 7), where:

S: Total water stored in the SVS land-surface scheme for a given date, in mm or kg/m^2.

$WVEG, SWE, WSL_x$: The three different SVS variables related to water storage (see Table 4), in mm or kg/m^2.

H_i: Thickness of the i^{th} SVS soil layer, in mm (see Table 4).

For the gridded surface variables (i.e. all except streamflow), the grid has a 0.09 degree resolution (~10km) and a size of 191x143 grid cells, but the values are not provided outside of the aforementioned watersheds (see point 5.1 in "Methods"). Moreover, these gridded variables are all valid over the land fraction of a grid cell only, and not over other surface types like water. In other words, these variables correspond to outputs of the GEM-Hydro land-surface scheme SVS. Therefore, over pixels that are covered with 100% water in the domain, all of the gridded surface variables of this dataset have a value of 0.0, except soil moisture which has a value of 1.0 for these water pixels, for aesthetic purposes. However, note that the subgrid-scale lakes were removed from the GEM-Hydro setup (see point 4.1 of "Methods"), meaning that most of the grid cells of the region are assumed to be occupied at 100% by land cover in GEM-Hydro, except around grid cells and in grid cells that are occupied by 100% water (see the field WT mentioned below). In any case, the gridded surface variables of this dataset can still be used as-is, for example to drive a routing model over the region (see Usage Notes). More information about the reason for (and the impact of) neglecting the subgrid-scale lakes in the GEM-Hydro setup used here can be found in point 4.1 of the "Methods" section. The land cover fraction inside each grid cell does not evolve with time in this version of GEM-Hydro and is provided in this dataset as the "WT" variable.

The SM values shared in this dataset (see Table 4) correspond to liquid water only, because the version of SVS used during this study does not include the representation of soil freeze/thaw processes, and therefore does not simulate the conversion from liquid to frozen soil water, during cold seasons. This is because when activating the soil freeze-thaw processes with SVS, spring freshets are generally strongly overestimated, partly because the model does not allow for infiltration into frozen ground, which could occur in reality because of macropores (Mohammed et al., 2018[38]). Work is under way to improve the representation of the soil freeze-thaw processes with SVS, and their impact on resulting streamflow simulations.

Table 4 summarizes the different files shared in this dataset, as well as their content. Note that each file of this dataset contains a unique hydrologic variable, over the region of interest and over the full period of interest, except for streamflow, where the file shared is a compressed file. Once unzipped, there will be one text file for each of the 212 flow gauges considered in the domain, containing the daily observed and simulated streamflow over the full period.

Table 4: list of files shared in this dataset and description of their content. See point 5.1 in "Methods" for more details about the computations performed with the actual GEM-Hydro outputs to produce the variables mentioned in this table.

Filename (Hydrologic variable)	Hydrologic Variable	Units	Definition
GEM-Hydro_calibrated_PR.nc	PR	mm/h	Hourly total precipitation over the hour preceding the date mentioned
GEM-Hydro_calibrated_ACWF.nc	ACWF	mm/h	Hourly evapo-transpiration over land surface over the hour preceding the date mentioned
GEM-Hydro_calibrated_TRAF.nc	TRAF	mm/h	Hourly surface runoff over land surface over the hour preceding the date mentioned
GEM-Hydro_calibrated_ALAT.nc	ALAT	mm/h	Hourly total lateral flow from land surface soil column (from all active soil layers) over the hour preceding the date mentioned
GEM-Hydro_calibrated_O1.nc	O1	mm/h	Hourly drainage from (vertical water flux leaving the) land surface last active soil layer over the hour preceding the date mentioned
GEM-Hydro_calibrated_WVEG.nc	WVEG	kg/m^2 or mm	Water stored in/on the land surface vegetation at the date mentioned, but weighted such that it corresponds to an average height of water valid over the whole land surface
GEM-Hydro_calibrated_SWE.nc	SWE	kg/m^2 or mm	Average Snow Water Equivalent over land surface at the date mentioned
GEM-Hydro_calibrated_WSL1.nc	WSL1	m^3/m^3	Soil moisture content for soil layer 1: for depth between 0 and 5 cm (volumetric fraction) at the date mentioned
GEM-Hydro_calibrated_WSL2.nc	WSL2	m^3/m^3	Soil moisture content for soil layer 2: for depth between 5 and 10 cm (volumetric fraction) at the date mentioned
GEM-Hydro_calibrated_WSL3.nc	WSL3	m^3/m^3	Soil moisture content for soil layer 3: for depth between 10 and 20 cm (volumetric fraction) at the date mentioned
GEM-Hydro_calibrated_WSL4.nc	WSL4	m^3/m^3	Soil moisture content for soil layer 4: for depth between 20 and 40 cm (volumetric fraction) at the date mentioned
GEM-Hydro_calibrated_WSL5.nc	WSL5	m^3/m^3	Soil moisture content for soil layer 5: for depth between 40 and 100 cm (volumetric fraction) at the date mentioned
GEM-Hydro_calibrated_WSL6.nc	WSL6	m^3/m^3	Soil moisture content for soil layer 6: for depth between 100 and 200 cm (volumetric fraction) at the date mentioned
GEM-Hydro_calibrated_WT.nc	WT	[-]	Fraction of grid cell occupied by land surface (constant over time).
GEM-Hydro_calibrated_streamflow.zip	Q	m^3/s	Pairs of observed and simulated mean daily streamflow at flow gauge locations in txt format. Once unzipped, one file per gauge. The .txt filenames correspond to a US or Canadian gauge ID.

Technical Validation

A) Streamflow performances

To perform the evaluation of GEM-Hydro streamflow simulations over the Great-Lakes and Ottawa river basins, the mean daily GEM-Hydro simulated streamflows were computed at the location of the 212 streamflow gauges used in the GRIP-GL project (Mai et al. 2022[12]), for which mean observed daily streamflow is available, over the period from January 1st 2001 to December 31st, 2017. Despite the GEM-Hydro simulations performed here also cover the year 2018, it was not considered in this evaluation, because the observations were gathered from the data of the GRIP-GL project, which does not include the year 2018. For a description of the flow gauges main attributes (gauge ID, river name, drainage area, mean elevation and mean annual runoff, etc.), please refer to the supplementary material of GRIP-GL (Mai et al. 2022[12]). The subbasins corresponding to these 212 flow gauges are shown on Figure 2. Figures 5 and 6 show boxplots of streamflow performances across these 212 flow gauges, either over the calibration (2008/01/01-2017/12/31) or validation period (2001/01/01-2007/12/31), for MESH-SVS-Raven and different versions of GEM-Hydro, and for three different scores. The scores include the revised KGE criteria (see Kling et al. 2012[28]), the Nash-Sutcliffe criteria (NSE, see Nash and Sutcliffe, 1970[39]) and a relative percent bias criteria (see Equation 8 below).

$$PBIAS = \frac{\sum_{i=1}^{N}(O_i - S_i)}{\sum_{i=1}^{N} O_i} * 100 \qquad \text{Eq. 8, where:}$$

$PBIAS$: Relative Bias in percent

O_i: Observed streamflow for day *i*

S_i: Simulated streamflow for day *i*

Therefore, when the $PBIAS$ value is positive (negative), it denotes an underestimation (overestimation) of the observed flows by the simulations.

First, it can be seen on Figure 5a that when the SVS and some routing parameters are calibrated with MESH-SVS-Raven and then transferred into GEM-Hydro, the streamflow performances remain relatively similar, highlighting the relevance of the approach used here to calibrate SVS and some routing parameters using another system than GEM-Hydro. The same is true when modifications are made to the GEM-Hydro setup in order to use a configuration that is more representative of the one used in the NSRPS system. This is encouraging because it supports the idea that the calibrated parameters obtained here with MESH-SVS-Raven were not overfitted to the specific setup configuration used with MESH-SVS-Raven and can actually be transferred to a GEM-Hydro version using a different setup configuration. The absence of over-calibration or overfitting in the calibrated parameter values obtained here is moreover supported by the evaluation of the auxiliary hydrologic and near-surface variables presented further down.

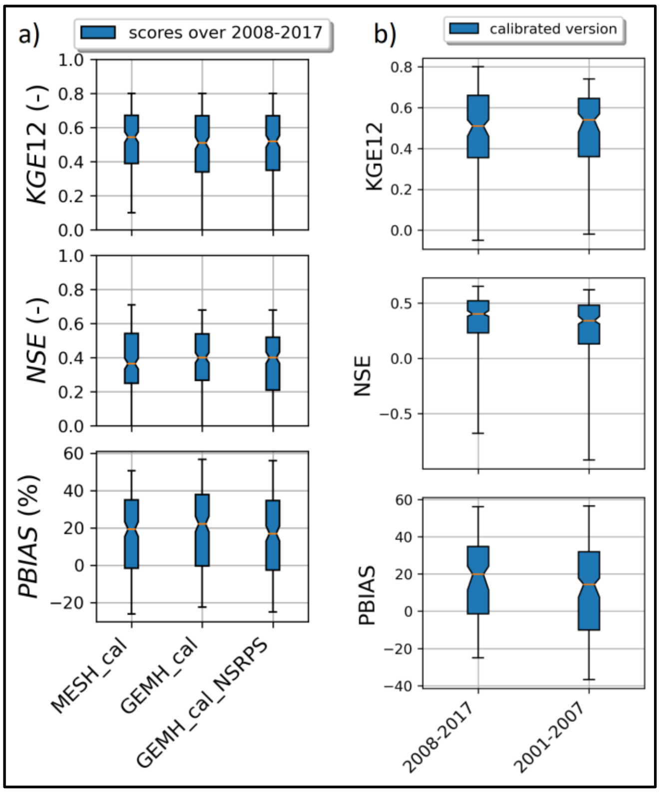

Figure 5: Boxplots of streamflow performances across the 212 flow gauges considered here, for the different calibrated models. a) performances over the calibration period (1st Jan. 2008- 31 Dec. 2017) for the three different calibrated model runs mentioned in this document. MESH_cal corresponds to the calibrated version of MESH-SVS-Raven using the GRIP-GL setup configuration, GEMH_cal corresponds to the calibrated version of GEM-Hydro using the GRIP-GL setup configuration, and GEMH_cal_NSRPS corresponds to the calibrated version of GEM-Hydro using the NSRPS setup configuration (see Methods and text for more details). b) performances for the calibrated version of GEM-Hydro using the NSRPS configuration over the calibration (20080101-20171231) and validation (20010101-20071231) periods. The target value for KGE12 (revised KGE) and NSE criteria is 1.0, while the

target value for PBIAS (relative percent bias) is 0.0. On figures 5 and 6, the lower and upper limits of the boxes correspond respectively to quartiles 25 and 75% of the 212 gauges' performances, the median score value is shown with a horizontal line in the thinner part of the box, and the lower and upper whiskers correspond respectively to percentiles 5 and 95% of the gauges' performances, while the outliers are not shown.

It can be seen on Figure 5b that the calibrated version of GEM-Hydro displays a strong temporal robustness, because the streamflow performances are generally similar between the calibration and validation periods. Finally, it can be seen on Figure 6 that the calibrated version of GEM-Hydro generally displays better streamflow performances than the default version of the model (without calibration), with a median KGE improvement close to 0.2 (see Figure 6b). It can also be seen on Figure 6b that regarding the KGE and NSE criteria, an improvement occurred for 75% of the 212 flow gauges considered. However, it also means that for 25% of the stations, the calibrated version of GEM-Hydro actually degrades flow performances compared to the default version of the model. When looking at flow hydrographs, this was generally attributed to overestimated streamflow peaks and flow volumes. This is supposed to be caused by the calibrated parameters obtained here, which imply a significant increase in the horizontal and vertical hydraulic conductivities in agricultural areas, especially for the watersheds of Lake Erie and Lake Ontario. This is because GEM-Hydro is missing an explicit representation of the tile drains in these areas, and because of the strategy employed here to represent the effect of these tile drains in the model (see Methods). Indeed, with this strategy, tile drains are assumed to be present in 100% of the agricultural areas, which is probably not the case in reality. Therefore, where tile drains are present in reality, the calibration methodology employed here generally leads to an improvement of the simulations compared to the default version of the model (but still to generally underestimated flow volumes, see for example Figure 7), while when tile drains may not be present in reality, the calibrated parameters lead to overestimated flows (see Figure 8). Therefore, the final calibrated values for GRKMO_A and KASMO_A (see Figure 3) consist of a trade-off between agricultural areas that are strongly impacted by human influence, and those that are less influenced. This may be improved in the future for example by using a GRKMO_A calibration parameter that would only be tied to areas where tile drains are actually present, and a KASMO_A calibration parameter only tied to areas where significant ploughing practices occur, instead of being applied to all agricultural areas. However, this information would need to be available on maps (over the US and Canada), which was not the case at the time of this study, to the extent of our knowledge. This is therefore dedicated to future work. For the time being, we consider that the calibrated version of GEM-Hydro, whose outputs are described here, generally represents an improvement upon its default counterpart, for most of the Great-Lakes and Ottawa River watersheds.

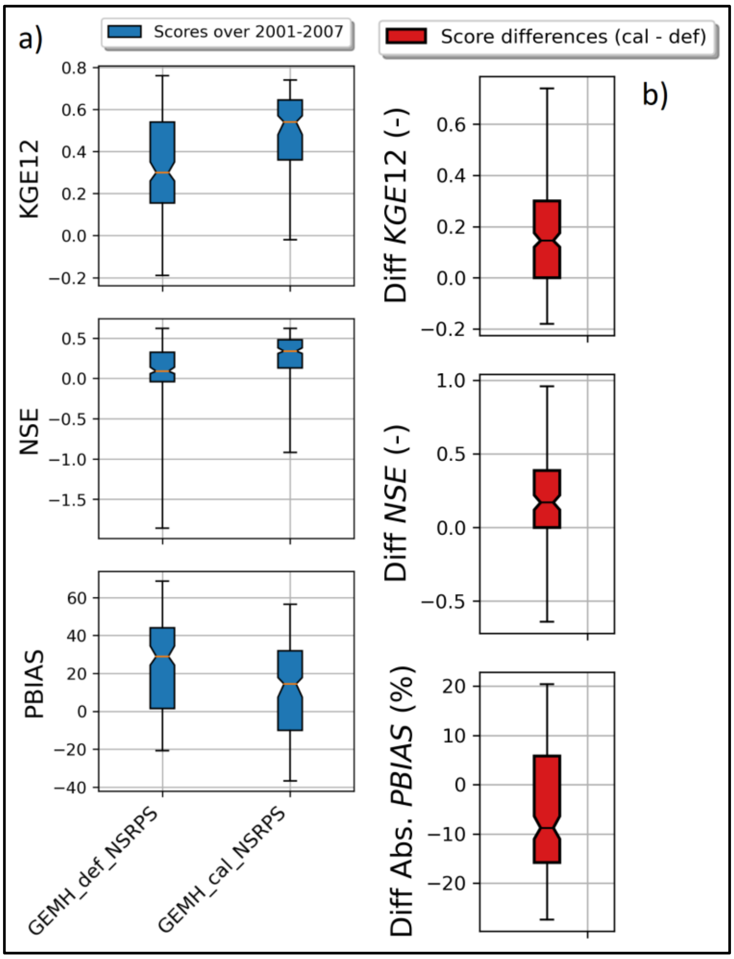

Figure 6: boxplots of streamflow performances across the 212 flow gauges considered here, for the default and calibrated versions of GEM-Hydro over the validation period (20010101-20071231), and using the NSRPS configuration. a) Performances of each version separately. GEMH_def_NSRPS corresponds to the default GEM-Hydro version. GEMH_cal_NSRPS corresponds to the calibrated version of GEM-Hydro. b) Differences in streamflow performances between the two versions of GEM-Hydro. For the KGE and NSE criteria, the difference between the calibrated and default versions of GEM-Hydro was computed at each gauge location, and values above 0.0 indicate an improvement of the calibrated upon the default version of the model. For the PBIAS criteria, the difference between the calibrated and default version of the model was computed using the absolute PBIAS value of each version and at each gauge location, and negative values indicate an improvement of the calibrated upon the default version of the model. See legend of Figure 5 for more details on the scores and the boxplots shown here.

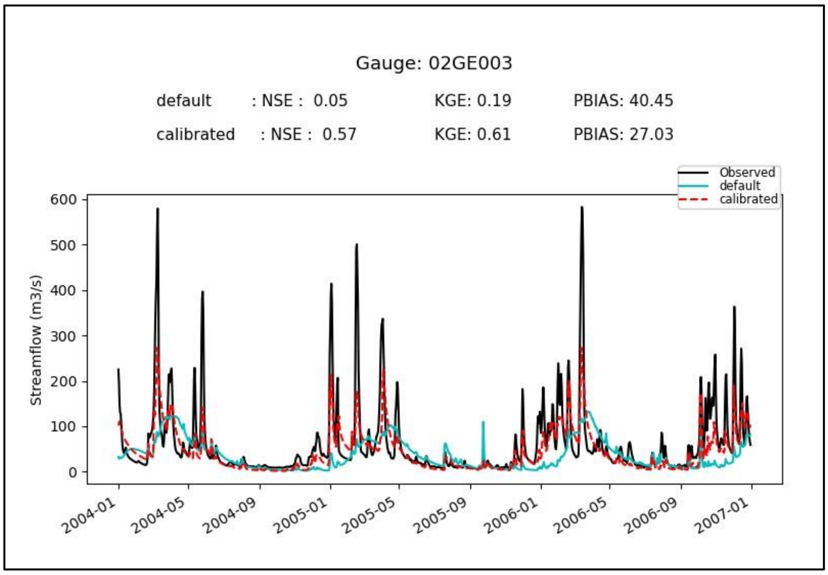

Figure 7: Hydrographs for gauge 02GE003, for the default and calibrated versions of GEM-Hydro, over a three-year period of the validation period. A three-year period is used here to better display the change in flow dynamics between the two versions. The flow observations are shown with the black line. The station 02GE003 is located on the northern shore of Lake Erie, in an agricultural area, and has a drainage area of 4498 km^2.

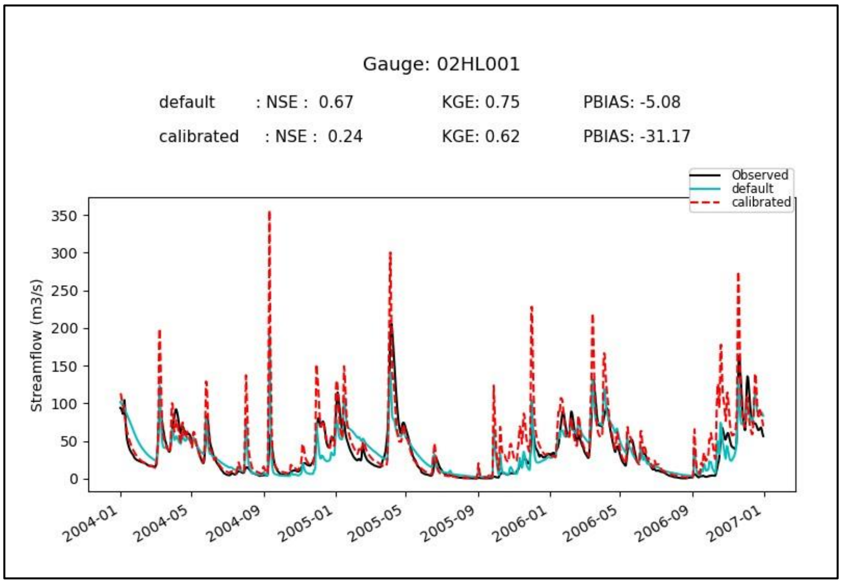

Figure 8: Hydrographs for gauge 02HL001, and for the default and calibrated versions of GEM-Hydro, over a three-year period of the validation period. A three-year period is used here to better display the change in flow dynamics between the two versions. The flow observations are shown with the black line. The station 02HL001 is located on the northern shore of Lake Ontario, in an agricultural area, and has a drainage area of 2673 km^2.

B) GEM-Hydro auxiliary Hydrologic variables

In order to evaluate the performances of the calibrated version of GEM-Hydro beyond streamflow performances, the default and calibrated versions of GEM-Hydro (using the final NSRPS configuration) were compared with regard to the performances of auxiliary hydrologic variables, as was done in GRIP-GL (Mai et al. 2022[12]), but also from the view point of near-surface meteorological variables (see point C below). This is important to evaluate a physically-based model with other variables than streamflow, especially in the case where the model was calibrated to maximize only streamflow performances, in order to make sure that the calibration process did not result in degrading other physical processes in the model (Kirchner 2006[40]). This can often happen during calibration, that the final performances are good with regard to streamflow but imply unrealistic physical processes. This is in opposition to the idea of "getting the right answer for the right reasons" (Kirchner 2006[40]). Getting the right answer for wrong reasons is linked with the notion of parameter equifinality (Beven and Binley 1992[41]): very different parameter sets, all leading to similar streamflow performances, can imply very different internal physical processes, some of which may be unrealistic to various degrees. This is especially important to assess the performances of the calibrated SVS version for other variables, in the context where this land-surface scheme could ultimately be two-way coupled with an atmospheric model, as mentioned in the "Methods" section.

Therefore, the goal here is to make sure that the performances of the calibrated version of GEM-Hydro remain similar or are better than the default version of the model, when looking at other outputs of the model. Here we focus on three hydrologic variables other than streamflow: the model total ET over land (which in GEM-Hydro is the sum of bare ground evaporation and vegetation ET, in mm), the SSM (mean soil moisture between 0 and 10cm depth, in m^3/m^3), and the mean SWE on the ground (in mm). In order to evaluate these variables, their mean daily value was computed from the hourly GEM-hydro outputs (or the total value in the case of AET), and compared to a reference dataset for each of them. The total AET over land is a direct output of GEM-Hydro that is being shared in this dataset, the mean SVS SWE was computed as specified in the Data Records section, and the SSM between 0 and 10cm was simply computed by taking the average of soil moisture for the first two SVS soil layers.

The reference datasets are mentioned in Table 5. They are the same as those used during GRIP-GL (Mai et al. 2022[12]). Note that these datasets do not consist of purely observed data, and as such, they cannot be considered as the "truth" for these variables. However, the idea here is not to perform a comprehensive evaluation of the calibrated version of GEM-Hydro, but rather to compare it to the default version of the model to make sure that the performances remain at least similar between the two. Note that regarding ERA-5 Land that is used as the reference for the SWE, it was preferred upon in-situ measurements of SWE because in-situ SWE measurements are not gridded (they are not available for all pixels of the region), and are not available on a daily basis either. However, an evaluation of the calibrated GEM-Hydro version and ERA-5 Land against in-situ SWE measurements is also performed here.

Table 5: List of the reference datasets used to evaluate the GEM-Hydro auxiliary hydrologic variables.

Variable	**Reference dataset**
Evapo-transpiration, in mm	**GLEAM v3.5b:** **GLEAM | Global Land Evaporation Amsterdam Model** **See also Martens et al. (2017[44])**
Soil moisture (0-10cm), in m^3/m^3	**GLEAM v3.5b:** **GLEAM | Global Land Evaporation Amsterdam Model** **See also Martens et al. (2017[44])**
SWE (mm)	**ERA-5 Land** **See product documentation and Muñoz Sabater (2019[45])**

The evaluation against these reference datasets was performed by computing the KGE criteria for each grid cell of the full domain, using the time-series of mean daily (or total daily for AET) values of these variables. Note that in this case, the KGE values correspond to the original (not revised) formulation of the KGE, as proposed by Gupta et al. (2009[42]). For AET and SSM, the period from 2003 to 2017 was used because this is the period for which the GLEAM v3.5 data were available, but for the comparison against ERA-5 Land regarding SWE, we used the period from 2001 to 2017. Finally, the GEM-Hydro outputs were re-gridded to the resolution of the reference datasets, which is of 0.25° for the GLEAM dataset (AET and SSM), and of 0.1° for ERA-5 Land (SWE). See Mai et al. (2022[12]) for more details on the auxiliary variables' evaluation.

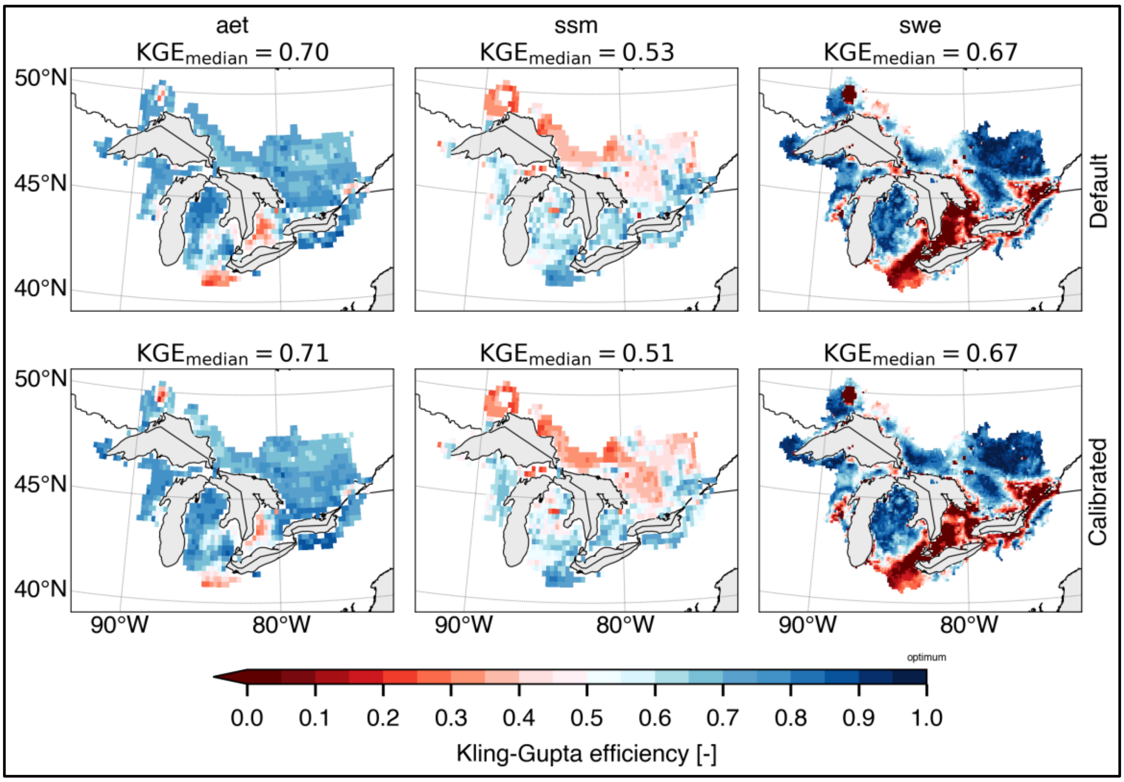

Figure 9: Comparison of the two versions of GEM-Hydro against reference datasets for the auxiliary hydrologic variables. The top row shows the comparison for the default GEM-Hydro version, while the comparison of the calibrated version is shown on the bottom row). The three auxiliary hydrologic variables considered here consist of evapo-transpiration (aet, left column), superficial soil moisture (ssm, middle column), and snow water equivalent (swe, right column).

It can be noticed on Figure 9 that the default and calibrated versions of GEM-Hydro generally display similar performances when compared to the reference datasets used here. Note that to achieve these results, a total of 6 full calibration trials were performed (see Methods), and several modifications to the calibration methodology were applied after each iteration because for the former trials, a degradation was sometimes noticed with the calibrated version of GEM-Hydro, for some of these auxiliary variables and in some regions. As previously mentioned, this is because of the equifinality issue, that especially arises when only calibrating a physical model to streamflow. However, note that performing multi-objective calibration to include for example these three auxiliary variables in the objective function during calibration is, on one hand, not a trivial task that can itself call for several iterations of the procedure (see Mai 2023[7]), and on the second hand, a methodology that could lead to suboptimal streamflow performances (Mei et al. 2023[9]). However, maximizing streamflow performances was the main objective of the calibration work performed here, in order to ultimately improve the NSRPS real-time streamflow forecasts performed at ECCC. This is why it was preferred to only target flow performances during calibration but evaluate the model based on other variables afterwards.

It can be noticed on Figure 9 that for each of the three GEM-Hydro auxiliary variables, there are some areas that display a stronger discrepancy against the reference data used here. For example, for AET, the areas with stronger discrepancy consist of the northern shore of Lake Erie, and the southern west portion of Lake Erie. In these areas, both GEM-Hydro versions tend to overestimate AET, which was diagnosed by looking at the bias component of the KGE (not shown here). Despite the exact source of this overestimation is not known, it was noticed that a better match between the calibrated GEM-Hydro version and the reference dataset was obtained in these areas for AET when using much higher values (close to 100.0) for the KASMO_A parameter (see Table 1), but to the detriment of SSM performances, however. Since it was not sure that this model response was due to the right reasons, a maximum value of 5.0 for KASMO_A was preferred (Table 1), leading to the results shown on Figure 9. Another issue in these areas is that the SVS model currently does not represent irrigation, which may have a strong effect on SSM.

Regarding SSM, the areas with the strongest discrepancies between GEM-Hydro and the reference dataset consist of the northern part of the Great-Lakes region. Again, while the reason is not exactly known, it has to be emphasized that these areas correspond to regions with a high fraction of high vegetation covers, where the remotely sensed satellite soil moisture data (that GLEAM assimilates) are known to be less accurate (Tong et al. 2020[43]). Therefore, it is not sure that GEM-Hydro actually performs worst in this northern region, with regard to SSM. It is also emphasized that the actual KGE performances are displayed here for SSM, while in the GRIP-GL project, only the correlation component of the KGE was shown for this variable, given that many models were not simulating the actual soil moisture variable, or not for the same depth. However, it is much easier to achieve satisfactory SSM simulations with regard to correlation only, rather than for the actual KGE value that is used here. Finally, regarding SWE, there are several areas showing a strong discrepancy between GEM-Hydro and ERA-5 Land.

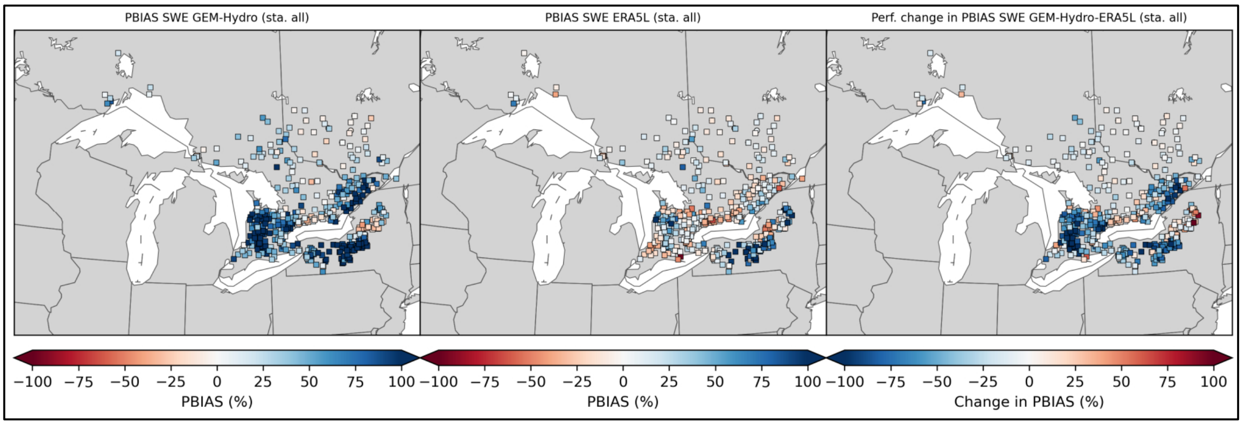

Figure 10: Comparison between the bias of SWE simulations from GEM-Hydro and from ERA-5 Land, over the watersheds considered in this study and over the period from October 1st 2001 to October 1st 2018 (17 winters). The observations mainly consist of in-situ manual snow surveys included in the CanSWE database (Vionnet et al. 2021[46]) and taken from the northeastern US databases (Mortimer et al., 2022[47]). Only stations with at least 20 observations (~one average per winter) available were considered. For the left (GEM-Hydro bias) and middle (ERA-5 Land bias) panels, the bias shown consists of the relative bias expressed in % (see Equation 8), but using the difference between the simulated and observed values, in opposition to Equation 8 (see text for more details). Blue (red) colors imply an over- (under-)estimation of the observed values. The differences between the absolute PBIAS values of ERA-5 Land and GEM-Hydro are shown on the right panel, such that red (blue) colors indicate that GEM-Hydro SWE simulations have less (more) bias than those of ERA-5 Land.

For the areas exhibiting strong differences between ERA-5 Land and GEM-Hydro in terms of SWE simulations (see Figure 9), it can be seen on Figure 10 that ERA-5 Land is generally better than GEM-Hydro. The strong positive *PBIAS* values observed for GEM-Hydro SWE simulations in some areas of Figure 10 correspond well to the areas for which strong differences were noticed between GEM-Hydro and ERA-5 Land, in terms of SWE (see Figure 9). This SWE overestimation by GEM-Hydro (see legend of Figure 10) in these areas was less pronounced (but still present) when using the traditional 0°-threshold method to separate liquid and solid precipitation in GEM-Hydro (not shown), instead of the Harder and Pomeroy (2013[30]) method used here for precipitation-phase partitioning (see Methods). However, the evaluation of the 0°-threshold method has shown that it creates a strong understimation of snowfall occurrence, for many regions of Canada (see for example Vionnet et al. 2022[31]). On the other hand, evaluation of the precipitation-phase partitioning method from Harder and Pomeroy (2013[30]) over the Great Lakes has shown that it can lead to an overestimation of snowfall occurrence (not shown here). Therefore, it is supposed that the positive bias noticed here with GEM-Hydro with regard to SWE for some areas, results from a combined positive bias in CaSR v2.1 winter precipitation forcings (Gasset et al. 2021[24]) and a positive bias in snowfall occurrence from Harder and Pomeroy (2013[30]). Additional investigations are required in the context of the preparation of the CaSR v3.0 and are beyond the scope of this document. However, there are areas for which the GEM-Hydro simulations performed here are very competitive with (or very close to) ERA-5 Land with regard to SWE simulations (see Figures 9 and 10).

C) GEM-Hydro near-surface meteorological variables

As explained in the "Methods" section, it is also important to make sure that the calibrated version of GEM-Hydro does not degrade the near-surface variables simulated by the model, as compared to the default version. This is related to the fact that ultimately, the calibrated version of the surface component of GEM-Hydro could be two-way coupled with ECCC atmospheric models. The three surface variables considered here consist of 2-m air temperature (TT, in °C.), 2-m dew point (TD, in °C.), and 10-m wind speed (UV, in $m.s^{-1}$). It is possible to evaluate these GEM-Hydro variables because they are simulated by the model and do not consist of forcing variables. Indeed, GEM-Hydro is driven with atmospheric forcings (like air temperature and humidity and wind) corresponding to the lowest prognostic level of an atmospheric model, which in the case of the CaSR v2.1, corresponds approximately to 40 m. In order to perform the evaluation of the default and calibrated versions of GEM-Hydro with regard to these variables, the ECCC internal verification tool "EMET" was used. The evaluation was performed by considering in-situ observations from the METAR, SYNOP, and SWOB observation networks over the full domain considered here, using the hourly GEM-Hydro outputs, and over the period from 2013 to 2017 included. Each one of the two GEM-Hydro versions was evaluated by computing the mean bias and the standard deviation of the error of the simulations, as compared to the observed values of a given variable, over a given period. The average of a given score was then computed over the area of interest. Then, the performances of the two versions were compared and are shown in Table 5, which shows the relative differences between the performances of the two versions of GEM-Hydro (i.e., default and calibrated).

To compute the relative bias differences, Equation 9 below was used:

$$REL.\Delta BIAS = \frac{|S2|-|S1|}{|S2|} * 100 ,$$ Eq. 9, where:

$REL.\Delta BIAS$: Relative BIAS difference (in %)

$S2$: Mean Bias of the default version of GEM-Hydro

$S1$: Mean Bias of the calibrated version of GEM-Hydro

The equation of the standard deviation of the error is given below.

$$STD = \sqrt{\frac{1}{N} * \sum_{i=1}^{N}(X_i - \bar{X})^2}, \qquad \text{Eq. 10, where:}$$

STD: Standard deviation of the error

N: total number of observation-simulation pairs.

$X_i = P_i - O_i$, where P_i is the simulated (or forecasted) value for time-step *i*, and O_i is the observed value for time-step *i*.

As such, the standard deviation of the error can be seen as a measure of the variations of the model errors from which the mean bias would have been removed.

To compute the relative difference of the standard deviation of the error, Equation 11 below was used.

$$REL.\Delta STDEV = \frac{S2 - S1}{S2} * 100, \qquad \text{Eq. 11, where:}$$

$REL.\Delta STDEV$: Relative difference of the standard deviation of the error (in %).

$S2$: Standard deviation of the error of the default GEM-Hydro version

$S1$: Standard deviation of the error of the calibrated GEM-Hydro version

Table 6: Comparison between the default and calibrated versions of GEM-Hydro with regard to surface variables. TD: 2-m dew point temperature. TT: 2-m air temperature. UV: 10-m wind speed. Relative differences of the bias and the standard deviation of the error between the two experiments are shown. Warm colors and positive values denote an improvement of the calibrated version upon the default one, while cold colors and negative values denote a degradation. Note that the relative differences are shown here for different periods. For the values split by season, note that for each season, the average of the values for the years 2016 and 2017 is shown. See text for more details on the computation of the relative differences shown here.

	Period / variable	FULL	WINTER	SPRING	SUMMER	FALL
		20130101 / 20171231	0101-0331	0401-0630	0701-0930	1001-1231
REL. ΔBIAS	TD	1.45%	1.06%	2.50%	1.30%	1.02%
	TT	1.00%	-0.28%	1.40%	0.37%	0.15%
	UV	-0.27%	0.12%	-0.49%	-0.41%	-0.17%
REL. ΔSTDEV	TD	0.08%	-0.03%	0.13%	0.04%	0.01%
	TT	-0.05%	-0.03%	-0.06%	0.02%	0.01%
	UV	-0.02%	-0.01%	-0.04%	-0.04%	-0.02%

It can be seen from Table 6 that the differences with regard to near-surface variables, between the two versions of GEM-Hydro, are generally small, and can be considered neutral in terms of the standard deviation of the error. A generally small improvement of the calibrated upon the default version can be noticed for TT and TD Bias, while small degradations are noticed for the wind speed. Note that regarding TT, the conclusions actually depend on the season considered. Regarding wind speed, it is not exactly known why a small degradation could occur with the calibrated version, especially when considering that the multiplying coefficient related to surface roughness was not employed in this final calibration framework (see Methods). However, the UV bias differences are generally very small between the two versions, of the order of 0.015 m.s^{-1} (not shown here), which can be considered negligible. The same is true with regard to differences related to TT and TD. For example, the best improvement, which was noticed for the spring season and for TD Bias (Table 6), involves differences between the two GEM-Hydro versions of the order of 0.05 °C. maximum (not shown here), which can also be considered negligible. Note, however, that the model has a tendency to overestimate TD by 1 to 2 °C during the day (local time corresponds to UTC -4 hours during the spring period), over this region and for the spring season of 2017. However, when looking at the TD bias evolution as a function of the hour of the day but when considering the full period from January 1st 2013 to December 31st, 2017, this TD overestimation reaches 1 °C. maximum (not shown). Regarding 2-m air temperature (TT), GEM-Hydro simulations however have a tendency to underestimate this variable by about 0.5 °C. during the night (not shown).

Usage notes

It is emphasized here that the gridded surface variables shared in this dataset correspond to outputs of the SVS model, which is the land surface scheme of the GEM-Hydro model. Therefore, these surface variables are valid over land only, and not over other types of continental surfaces, such as glaciers, water, or ice. However, the water (and ice) surfaces were neglected in the GEM-Hydro setup employed here, implying that most grid-cells of the region of interest (except inside big lakes such as the Great Lakes, for example) were assumed to be filled at 100% with land surfaces only. See the Methods' section (point 4.1) and the Data records' section for more information about this modelling choice. Despite of this, the surface fluxes shared in this dataset (including surface runoff, soil lateral flow, and soil base drainage) can still be used as inputs to any routing model implemented over a basin that is included in the region of interest, provided that this basin does not include grid-cells that are filled at 100% with water surfaces: otherwise, the routing model will miss the fluxes coming from these 100% water pixels. This can be ensured based on the "WT" variable shared in this dataset, which represents the percentage of the land surface that was considered inside each grid-cell (see Data Records), for the GEM-Hydro simulations performed here.

In order to drive a routing model with the surface fluxes shared in this dataset, the sum of surface runoff and lateral flow (i.e. the "TRAF" and "ALAT" variables of this dataset) should be directly given as inputs to the surface network of the routing model (i.e. lakes and rivers), while the SVS soil base drainage (the "O1" variable) should be provided first to a baseflow model (i.e., a model representing the aquifer), that is sometimes already included in the routing model. Indeed, the "O1" variable represents the aquifer recharge. The aquifer model will then simulate the baseflow that returns to the surface network of lakes and rivers, in the routing model. Note that the units of these surface fluxes correspond to kg/m^2/h or mm/h (assuming a density of 1000 kg/m^3, i.e. unsalted water). Therefore, when provided to a routing model, these fluxes should then be multiplied by a surface area (like the area of a subbasin or of the routing model grid-cell) in order to convert them into a volume of water per units of time. This is generally done in the routing model itself.

Finally, it is reminded here that the SM variables of this dataset (the WSL1-6 variables) only represent the liquid soil moisture content of a given soil layer. However, the version of SVS used in this work did not represent soil freeze-thaw processes (more information in the Data Records' section), such the WSL1-6 variables still represent the total water stored in the different soil layers, in the GEM-Hydro simulations performed during this work.

Code Availability

The SVS land-surface scheme and the WATROUTE routing scheme are both available in the MESH official repository available at this address: https://github.com/MESH-

Model/MESH-Releases (accessed on February 1st, 2024). The Raven routing model is open-source and can be accessed here: the Raven Hydrological Framework - Home Page (uwaterloo.ca) (accessed on November 30, 2023). The DDS calibration algorithm is available in the Ostrich calibration toolkit and can be accessed here: OSTRICH Optimization Software Toolkit (uwaterloo.ca) (accessed on November 30, 2023). The MESH-SVS-Raven setups used in this study to calibrate the SVS and routing parameters may be shared upon reasonable request. GEM-Surf (the surface component of GEM-Hydro) is open-source and is available on Github: https://github.com/ECCC-ASTD-MRD/sps.git GEM-Surf can be run outside of ECCC informatic infrastructure. However, it still needs to rely on forcing and geophysical fields in the “standard file” format (a binary file format only used internally at ECCC), and produces output files in this format as well. Some tools needed to manipulate and read files of this format are also available on github: https://github.com/ECCC-ASTD-MRD

Moreover, the WATROUTE version included in MESH cannot be run in a standalone mode, but only together with the SVS land-surface scheme included in MESH. The WATROUTE version used internally at ECCC cannot yet be run outside of ECCC infrastructure. Therefore, it is not yet possible to exactly replicate the GEM-Hydro simulations (I.e., by running GEM-Surf+WATROUTE) described here, outside of ECCC informatic infrastructure.

Acknowledgements

The authors wish to thank Dr. James Craig from the University of Waterloo, for his availability and support with regard to the use of the Raven routing model. The authors also wish to thank Dorothy Durnford from ECCC, for her strong involvement in the constant improvement of the WATROUTE model setups used at ECCC to simulate streamflow over Canada, including over the Great-Lakes and Ottawa River basins.

Author contributions

-Étienne Gaborit performed the calibration experiments with MESH-SVS-Raven, performed the GEM-Hydro simulations, processed the GEM-Hydro outputs, performed the streamflow evaluation and the near-surface variables’ evaluation, published the GEM-Hydro data on the Federated Research Data Repository (FRDR), and wrote the Data descriptor.

-Juliane Mai performed the evaluation of the GEM-Hydro auxiliary hydrologic variables for all of the calibration trials involved during this work (Figure 9), and actively participated in the preparation of the Raven setups used in this study, and in the collection of the observed streamflow time-series and the reference datasets used to evaluate the auxiliary hydrologic variables (Table 5), which were inherited from the GRIP-GL project (Mai et al. 2022[12]).

-Vincent Vionnet performed the comparison between GEM-Hydro and ERA-5 Land SWE biases (Figure 10), significantly contributed to improving the driver script used at ECCC to run GEM-Surf open-loop simulations, and performed a comprehensive review of this document.

-Dan Princz performed the integration of the SVS code into the MESH Platform, and is the main lead developer of the MESH platform. He developed the MESH functionalities required to be able to couple MESH with the Raven routing model, as well as the scripts needed to convert a GEM-Hydro setup into a MESH-SVS(-WATROUTE) setup.

-Hongren Shen was the main person responsible for the preparation of the Raven routing model setups used in this work and that were inherited from the GRIP-GL project (Mai et al. 2022[12]).

-Bryan Tolson was Hongren Shen and Juliane Mai’s supervisor at the time of the GRIP-GL project, and contributed significantly to many of the methodological choices made in the GRIP-GL project, which strongly influenced the methodology employed here.

-Vincent Fortin is Étienne Gaborit’s supervisor and also contributed significantly to many methodological choices made during the GRIP-GL project and the resulting calibration trials that lead to the dataset described in this document.

Competing interests

All of the authors of this data descriptor declare that they have not any conflict of interest with regard to the work and the data related to this document, or to the content of this document.

References

1. Gaborit, É., et al. A hydrological prediction system based on the SVS land-surface scheme: efficient calibration of GEM-Hydro for streamflow simulation over the Lake Ontario basin. *Hydrol. Earth Syst. Sci.*, **21(9)**, 4825-4839. https://doi.org/10.5194/hess-21-4825-2017 (2017)
2. Vionnet, V., et al. Assessing the factors governing the ability to predict late-spring flooding in cold-region mountain basins. *Hydrol. Earth Syst. Sci.*, **24(4),** 2141-2141 (2020).
3. Durnford, D., et al. Hydrological prediction systems at Environment and Climate Change Canada. *AMS 101st annual meeting*, January 9-15, 2021. Poster presentation (586) and extended abstract. Available at: https://ams.confex.com/ams/101ANNUAL/meetingapp.cgi/Paper/383559
4. Baroni, G., et al. A comprehensive distributed hydrological modeling intercomparison to support process representation and data collection strategies. *Water Resour. Res.*, ***55*(2),** 990-1010 (2019).
5. Budhathoki, S., Rokaya, P., and Lindenschmidt, K. E. Improved modelling of a Prairie catchment using a progressive two-stage calibration strategy with in-situ soil moisture and streamflow data. *Hydrol. Res.*, ***51*(3),** 505-520 (2020).
6. Hirpa, F. A., et al. Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *J. Hydrol.*, ***566*,** 595-606 (2018).
7. Mai, J. Ten strategies towards successful calibration of environmental models. *J. Hydrol.*, ***620*,** 129414 (2023).
8. Bajracharya, A. R., Ahmed, M. I., Stadnyk, T., and Asadzadeh, M. Process based calibration of a continental-scale hydrological model using soil moisture and streamflow data. *J. Hydrol.: Regional Studies*, ***47*,** 101391 (2023).
9. Mei, Y., et al. Can hydrological models benefit from using global soil moisture, evapotranspiration, and runoff products as calibration targets? *Water Resour. Res.*, ***59*(2)**, e2022WR032064 (2023).
10. Demirel, M. C., et al. Tradeoffs between temporal and spatial pattern calibration and their impacts on robustness and transferability of hydrologic model parameters to ungauged basins. *Water Resour. Res.*, ***60*(1),** e2022WR034193 (2024).
11. Mai, J., et al. Great Lakes runoff intercomparison project phase 3: Lake Erie (GRIP-E). *J. Hydrol. Eng.*, ***26*(9),** 05021020. https://doi.org/10.1061/(ASCE)HE.1943-5584.0002097 (2021).
12. Mai, J., et al. The Great Lakes runoff intercomparison project phase 4: the Great Lakes (GRIP-GL). Hydrol. Earth Syst. Sci., **26,** 3537–3572. https://doi.org/10.5194/hess-26-3537-2022 (2022).
13. Gaborit, É. Hydrologic outputs from the 2023 calibrated version of GEM-Hydro over the Great-Lakes and Ottawa River region. Le Dépôt fédéré de données de recherche. Federated Research Data Repository (FRDR). https://doi.org/10.20383/103.0847 (2024).

14. Bernier, N. B., Bélair, S., Bilodeau, B., and Tong, L. Near-surface and land surface forecast system of the Vancouver 2010 winter Olympic and paralympic games. *J. Hydrometeorol.*, **12,** 508– 530. https://doi.org/10.1175/2011JHM1250.1 (2011).
15. Alavi, N., et al. Warm season evaluation of soil moisture prediction in the Soil, Vegetation and Snow (SVS) scheme. *J. Hydrometeorol.*, **17,** 2315–2332, https://doi.org/10.1175/JHM-D-15-0189.1 (2016).
16. Husain, S. Z., et al. The multi-budget Soil, Vegetation, and Snow (SVS) scheme for land surface parameterization: offline warm season evaluation. *J. Hydrometeorol.*, **17,** 2293–2313, https://doi.org/10.1175/JHM-D-15-0228.1 (2016).
17. Leonardini, G., et al. Evaluation of the Soil, Vegetation, and Snow (SVS) land surface model for the simulation of surface energy fluxes and soil moisture under snow-free conditions. *Atmosphere*, **11(3)**, 278, https://doi.org/10.3390/atmos11030278 (2020).
18. Leonardini, G., F. Anctil, V. Vionnet, M. Abrahamowicz, D. F. Nadeau, and V. Fortin. Evaluation of the snow cover in the Soil, Vegetation, and Snow (SVS) land surface model. *J. Hydrometeor.*, **22,** 1663–1680, https://doi.org/10.1175/JHM-D-20-0249.1 (2021).
19. Kouwen, N. *WATFLOOD/WATROUTE Hydrological Model Routing & Flow Forecasting System*. Department of Civil Engineering, University of Waterloo, Waterloo, ON, 267 pp. 2010.
20. Yassin, F., Razavi, S., Elshamy, M., Davison, B., Sapriza-Azuri, G., and Wheater, H. Representation and improved parameterization of reservoir operation in hydrological and land-surface models. *Hydrol. Earth Syst. Sci.*, **23(9)**, 3735-3764 (2019).
21. Gaborit, É., Fortin, V., and Durnford, D. On the implementation of the dynamically zoned target release reservoir model in the GEM-Hydro streamflow forecasting system. *Can. J. Civ. Eng.*, **49**(10): 1582-1594, https://doi.org/10.1139/cjce-2021-0507, (2022).
22. Pietroniro, A., et al. Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. *Hydrol. Earth Syst. Sci.*, **11(4)**, 1279-1294 (2007).
23. Craig, J. R., et al. Flexible watershed simulation with the Raven hydrological modelling framework. *Environ. Model. Soft.*, **129,** 104728 (2020).
24. Gasset, N., et al. A 10 km North American precipitation and land-surface reanalysis based on the GEM atmospheric model. *Hydrol. Earth Syst. Sci.*, ***25*(9)**, 4917-4945 (2021).
25. Garnaud, C., et al. Improving snow analyses for hydrological forecasting at ECCC using satellite-derived data. *Remote Sens., **13***, 5022, https://doi.org/10.3390/rs13245022 (2021).
26. Gaborit, É., Ricard, S., Lachance-Cloutier, S., Anctil, F., and Turcotte, R. Comparing global and local calibration schemes from a differential split-sample test perspective. *Can. J. Earth Sci.*, **52(11)**, 990-999, https://doi.org/10.1139/cjes-2015-0015 (2015).

27. Tolson, B. A., and Shoemaker, C. A. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resour. Res.*, ***43*****(1)** (2007).
28. Kling, H., Fuchs, M., and Paulin, M. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.*, **424,** 264-277 (2012).
29. De Schepper, G., Therrien, R., Refsgaard, J. C., and Hansen, A. L. Simulating coupled surface and subsurface water flow in a tile-drained agricultural catchment. *J. Hydrol.*, ***521,*** 374-388 (2015).
30. Harder, P., and Pomeroy, J. Estimating precipitation phase using a psychrometric energy balance method. *Hydrol. Process.*, ***27*****(13)**, 1901-1914 (2013).
31. Vionnet, V., et al. Snow level from post-processing of atmospheric model improves snowfall estimate and snowpack prediction in mountains. *Water Resour. Res.*, ***58*****(12)**, e2021WR031778 (2022).
32. USGS: USGS EROS archive- Digital Elevation- Global Multi-resolution Terrain Elevation Data 2010 (GMTED 2010), https://doi.org/10.5066/F7J38R2N, (2010). Accessed: 2023-12-05.
33. ESA: Land Cover CCI vegetation cover, http://maps.elie.ucl.ac.be/CCI/viewer/download.php (2015), accessed: 2023-12-05.
34. Shangguan, W., Dai, Y., Duan, Q., Liu, B., and Yuan, H. A global soil dataset for earth system modeling. *J. Adv. Model. Earth Syst.*, **6,** 249–263, https://doi.org/10.1002/2013MS000293 (2014).
35. Natural Resources Canada: National hydro network—NHN—Geobase series, https://open.canada.ca/data/en/dataset/a4b190fe-e090-4e6d-881e-b87956c07977 (2020), accessed: 2021-03-03.
36. USGS: National Hydrography Dataset, https://www.usgs.gov/core-science-systems/ngp/national-hydrography/national-hydrography-dataset (2021), accessed: 2023-12-05.
37. HydroSHEDS: HydroSHEDS website, https://www.hydrosheds.org/ (2021), accessed: 2023-12-05
38. Mohammed, A. A., Kurylyk, B. L., Cey, E. E., and Hayashi, M. Snowmelt infiltration and macropore flow in frozen soils: Overview, knowledge gaps, and a conceptual framework. *Vadose Zone J.*, ***17*****(1),** 1-15 (2018).
39. Nash, J. E., and Sutcliffe, J. V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.*, **10(3)**, 282-290 (1970).
40. Kirchner, J. W. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.*, ***42*****(3)** (2006).
41. Beven, K., and Binley, A. The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process.*, ***6*****(3)**, 279-298 (1992).
42. Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol., **377(1-2),** 80-91 (2009).

43. Tong, C., et al. Soil moisture retrievals by combining passive microwave and optical data. *Remote Sens.* ***12*,** 3173, https://doi.org/10.3390/rs12193173 (2020).
44. Martens, B., et al. GLEAM v3: satellite-based land evaporation and root-zone soil moisture, *Geosci. Model Dev.*, **10,** 1903–1925 (2017).
45. Muñoz Sabater, J. ERA5-Land hourly data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), https://doi.org/10.24381/cds.e2161bac (2019). Accessed on 2023-12-05.
46. Vionnet, V., Mortimer, C., Brady, M., Arnal, L., and Brown, R. Canadian historical snow water equivalent dataset (CanSWE, 1928–2020). *Earth Syst. Sci. Data*, ***13*(9),** 4603-4619 (2021).
47. Mortimer, C., et al. Benchmarking algorithm changes to the snow CCI+ snow water equivalent product. *Remote Sensing of Environment*, ***274***, 112988, https://doi.org/10.1016/j.rse.2022.112988 (2022).