The Great Lakes Runoff Intercomparison Project Phase 4: The Great Lakes (GRIP-GL)

– Supplementary Material –

Juliane Mai[★,1], Hongren Shen[★,1], Bryan A. Tolson[★,1], Étienne Gaborit[★,2], Richard Arsenault[3], James R. Craig[1], Vincent Fortin[2], Lauren M. Fry[4], Martin Gauch[5], Daniel Klotz[5], Frederik Kratzert[5,6], Nicole O'Brien[7], Daniel G. Princz[8], Sinan Rasiya Koya[9], Tirthankar Roy[9], Frank Seglenieks[7], Narayan K. Shrestha[7], André G. T. Temgoua[7], Vincent Vionnet[2], and Jonathan W. Waddell[10]

[★]Lead contributors. Other authors are ordered alphabetically by last name.
[1]Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada.
[2] Meteorological Research Division, Environment and Climate Change Canada, Dorval, QC, Canada.
[3]Department of Construction Engineering, École de technologie supérieure, Montreal, QC, Canada.
[4]Great Lakes Environmental Research Laboratory, National Oceanic and Atmospheric Administration, Ann Arbor, MI, USA.
[5]Institute for Machine Learning, Johannes Kepler University, Linz, Austria.
[6]Google Research, Vienna, Austria.
[7]National Hydrological Service, Environment and Climate Change Canada, Burlington, ON, Canada.
[8]National Hydrological Service, Environment and Climate Change Canada, Saskatoon, SK, Canada.
[9]Department of Civil and Environmental Engineering, University of Nebraska–Lincoln, Lincoln, NE, USA.
[10]Great Lakes Hydraulics and Hydrology Office, U.S. Army Corps of Engineers, Detroit, MI, USA.

Correspondence: Juliane Mai (juliane.mai@uwaterloo.ca)

S.1 Comparison of gridded ERA5-Land SWE estimates and ground-truth SWE observations from CanSWE database

The ERA5-Land SWE estimates (Muñoz Sabater, 2019) were compared with the SWE observations available in the version 2 of the CanSWE dataset (Vionnet et al., 2021). CanSWE combines historical manual (snow surveys) and automatic SWE measurements collected across Canada by different provincial and territorial agencies as well as hydro-power companies. CanSWE is an evolving dataset. Please refer to the following repository to check for version updates of the dataset: https://10.5281/zenodo.4734371.

In this study, only manual SWE measurements were used to make sure that the snow data would be available over the period from 2000 to 2017. In Ontario and Quebec, manual SWE data are collected on a biweekly to monthly basis. The times series of ERA5-Land SWE estimates were extracted at the grid cells corresponding to the snow observation locations available in CanSWE. Only stations with at least 34 observations (i.e., on average at least two observations per year during the study period from 2001 to 2017) were kept in the analysis. A total of 272 stations was considered and the results are shown in Fig. S1.
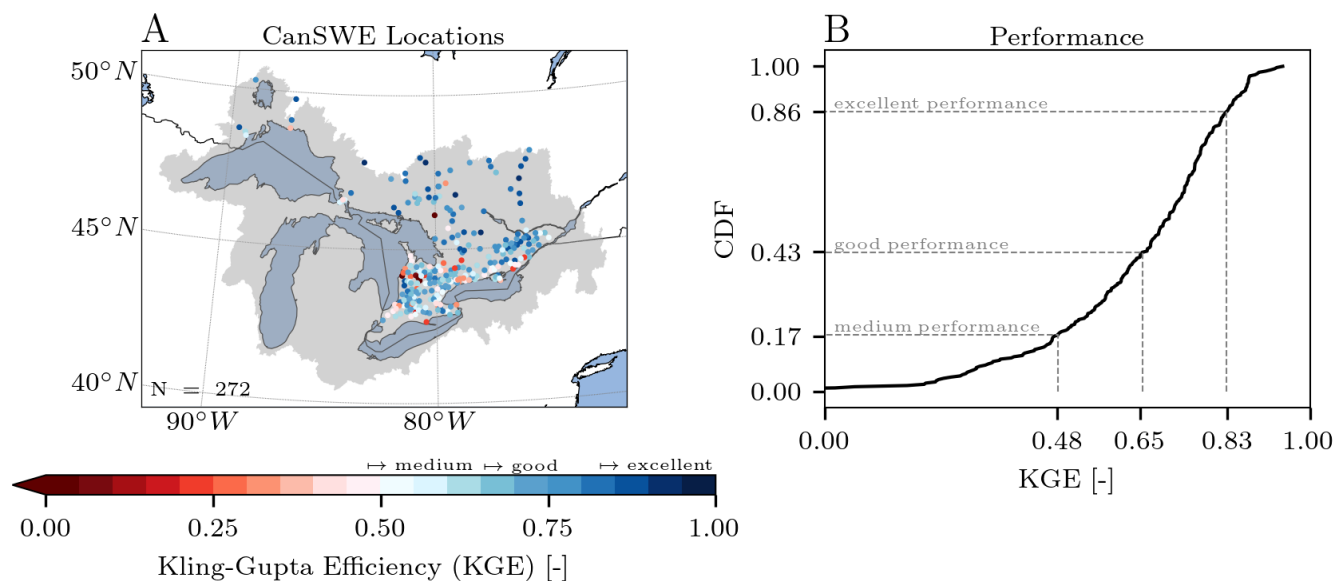
Figure S1. Comparison of ERA5-Land and CanSWE snow water equivalent estimates. Panel (A) displays the spatial distribution of the Kling-Gupta efficiency (KGE) between the CanSWE observations (obs) and ERA5-Land (sim) estimates at the $N = 272$ snow observation locations available in the CanSWE dataset with at least 34 observations (i.e., on average at least two observations per year during the study period from 2001 to 2017). Panel (B) summarizes these KGE values by a cumulative distribution function showing that around 83% of the stations have medium performance, 57% are good, and 14% show an excellent performance.

ERA5-Land shows good performances (KGE larger than 0.65) at 57% of the stations (Fig. S1B). These stations are mostly located in the Ottawa River watershed, in the Lake Superior watershed and in the northern part of the Lake Huron watershed (Fig. S1A). These regions are characterized by the largest mean winter SWE in the Great Lakes regions (Fig. 2 in the main

manuscript). More contrasted results, including stations with performances less than medium (KGE lower than 0.48), are found in the Canadian portions of the Lake Ontario and Lake Erie watersheds. There regions are characterized by milder winter conditions than the northern of the Great Lakes region with frequent occurrences of winter precipitation near 0°C (Mekis et al., 2020) associated challenging conditions for snowpack models due to complex precipitation phase and mid-winter melt resulting from rain-on snow events. Nonetheless, ERA5-Land remains a robust gridded SWE product in this region, especially when compared with other existing gridded SWE datasets such as the US SNODAS analyses characterized by a strong underestimation of SWE in Ontario (King et al., 2020).

The scripts to derive the estimates presented here are available on the GitHub associate with this project (https://github.com/julemai/GRIP-GL/tree/master/scripts/compare_ERA5-swe_CanSWE).

S.2   Additional information about participating models

In this section additional information will be provided about the participating models. Besides additional descriptions about model setup and calibration detail, a table of the parameters used for the model calibration is provided for each model. The section is organized as follows: globally calibrated Machine-Learning based LSTM-lumped model in Sect. S.2.1, the locally calibrated models LBRM-CC-lumped in Sect. S.2.2, HYMOD2-lumped in Sect. S.2.3, GR4J-lumped in Sect. S.2.4, HMETS-lumped in Sect. S.2.5, Blended-lumped in Sect. S.2.6, Blended-Raven in Sect. S.2.7, and VIC-Raven in Sect. S.2.8, as well as the regionally calibrated SWAT-Raven in Sect. S.2.9, WATFLOOD-Raven in Sect. S.2.10, MESH-CLASS-Raven in Sect. S.2.11, MESH-SVS-Raven in Sect. S.2.12, and GEM-Hydro-Watroute in Sect. S.2.13.

S.2.1   LSTM-lumped

The final LSTM model is an ensemble of 10 models with different random seeds. All models share the same architecture of 256 hidden states and use an input sequence length of 365. That is, a single day of discharge is predicted from a sequence of the previous 365 input time steps. We optimized the model parameters with an Adam optimizer (Kingma and Ba, 2015) on an loss function that approximates the NSE (Kratzert et al., 2019) for 20 epochs at a learning rate of 5×10^{-4} and another 10 epochs at a learning rate of 1×10^{-4} To regularize the model, we further applied dropout to the LSTM output with probability of 0.4 (Hinton et al., 2012). We stabilized the training procedure by clipping gradients to a maximum norm of 1 and initializing the LSTM forget gate with a bias of 3 (Gers et al., 1999).

The LSTM ingests the following input variables to predict daily discharge:

daily meteorological forcings:

– total precipitation,
– minimum and maximum temperature,
– mean downward solar flux,
– specific humidity,
– surface pressure,

3

- u/v components of wind, and

- potential evapotranspiration calculated following the Priestley-Taylor equation from the FAO-56 guidelines (Allen et al.,
  1998), using minimum and maximum temperature, solar radiation, latitude, elevation, and day of year as inputs and
  assuming that ground heat flux is zero at daily time steps (see also Newman et al., 2015)

static basin characteristics:

- in total 27 static input variables were used as inputs in the LSTM setup (Tab. S2)

- climatologic characteristics were calculated based on the calibration period (2000 to 2010)

- static features are fed into the LSTM at every time step alongside the meteorological forcings (see: Kratzert et al., 2019)

To ensure a model configuration that also works well in spatial validation, the LSTM hyperparameters (see Tab. S1) were tuned in a 5-fold cross-validation setting (Hastie et al., 2009) as follows: First, we randomly split the calibration basins into five partitions and trained each hyperparameter configuration on all combinations of four partitions. We repeated this process with three different random seeds and tested each setup on the corresponding remaining fifth partition. Second, we calculated the average KGE across all seeds for each basin in the partition that was not trained on and then took the median across all these basins. Finally, given the best hyperparameter configuration, we trained an ensemble of ten models with different random seeds on the full calibration period of all calibration basins combined and averaged their predictions.

The model was set up using the NeuralHydrology Python library in version 0.9.11-beta3 (Kratzert et al., 2022). The configuration files and scripts to reproduce the LSTM setup and training are available on the GitHub associated with this project (https://github.com/julemai/GRIP-GL/tree/master/scripts/MachineLearning).

Table S1: **The hyperparameters calibrated for the LSTM-lumped model.** The first three parameters are tested with the listed discrete values while the fourth parameter (epochs) was tested for a range of integer values in order to determine the best hyperparameter setting for this application. The best hyperparameter values found are highlighted in bold font. The parameter names are also the names of the NeuralHydrology configuration value. A brief description is given for each parameter.

Parameter	Parameter value/range	Description
learning_rate	{0: 0.001}, {0: 0.005, 10: 0.001, 20: 0.0005}, {0: 0.001, 20: 0.0005, 30: 0.0001}, **{0: 0.0005, 20: 0.0001, 30: 5e-05}**, {0: 0.0001, 20: 5e-05, 30: 1e-05}	Learning rate schedule of the optimizer. E.g., 0: x, 10:y means learning rate x for epochs 0 to 9, learning rate y after epoch 9
hidden_size	64, 128, **256**	Number of LSTM cell states
batch_size	**64**, 128, 256	Number of samples per mini-batch during gradient descent

Continued on next page

4

Parameter	Parameter value/range	Description
epochs	[3, 45]; **best: 30**	Number of training passes through the entire dataset (we evaluated every third epoch, so any value $3 \cdot k$ with $k \in [1, 15]$ was evaluated)

Table S2: **Static basin characteristic used for the LSTM-lumped model.** This lists the static basin characteristics exclusively derived from the common datasets (Sect. 2.2 of main manuscript) and used as input variables to the LSTM model. [†] Due to a data preprocessing mistake, only the average of the lower two soil layers from the soil dataset was considered for the soil related attributes.

Attribute name	Description
p_mean	Mean daily precipitation
pet_mean	Mean daily potential evapotranspiration
aridity	Ratio of mean PET to mean precipitation
t_mean	Mean daily temperature, i.e., (daily_min + daily_max) / 2
frac_snow	Fraction of precipitation falling on days with mean daily temperatures below 0°C
high_prec_freq	Fraction of high-precipitation days (≥ 5 times mean daily precipitation)
high_prec_dur	Average duration of high-precipitation events (number of consecutive days with ≥ 5 times mean daily precipitation)
low_prec_freq	Fraction of dry days (< 1 mm d^{-1} daily precipitation)
low_prec_dur	Average duration of dry periods (number of consecutive days with daily precipitation < 1 mm d^{-1})
mean_elev	Catchment mean elevation
std_elev	Standard deviation of catchment elevation
mean_slope	Catchment mean slope
std_slope	Standard deviation of catchment slope
area_km2	Catchment area
Temperate-or-sub-polar-needleleaf-forest	Fraction of land covered by "Temperate-or-sub-polar-needleleaf-forest"
Temperate-or-sub-polar-grassland	Fraction of land covered by "Temperate-or-sub-polar-grassland"
Wetland	Fraction of land covered by "Wetland"
Cropland	Fraction of land covered by "Cropland"

Attribute name	Description
Barren-Lands	Fraction of land covered by "Barren-Lands"
Urban-and-Built-up	Fraction of land covered by "Urban-and-Built-up"
Water	Fraction of land covered by "Water"
BD	Soil bulk density (g cm^{-3})[†]
CLAY	Soil clay content (% of weight)[†]
GRAV	Soil gravel content (% of volume)[†]
OC	Soil organic carbon (% of weight)[†]
SAND	Soil sand content (% of weight)[†]
SILT	Soil silt content (% of weight)[†]

S.2.2 LBRM-CC-lumped

The Large Basin Runoff Model (LBRM) (Crowley II, 1983) was developed for applications to historical simulation and seasonal forecasting of the Great Lakes water balance. As such, prior to this study, it has primarily been evaluated for simulating whole-basin runoff into the lakes. It is a lumped conceptual model that simulates the propagation of rainfall to watershed outflow using a series of cascading tanks, with transfers between tanks represented through a mass balance coupled with linear reservoir concepts. Calibrated model parameters include linear and partial linear reservoir coefficients to represent transfer between the cascading tanks, as well as a base temperature parameter that controls the potential evapotranspiration and an upper soil zone capacity parameter. Since the initial development of LBRM, updates have included the integration of an approach to approximate the Clausius-Clapeyron to mitigate the oversensitivity of evapotranspiration to temperature identified by Lofgren et al. (2011). This update, referred to as LBRM-CC-lumped in this manuscript, demonstrated by Lofgren and Rouhana (2016) and implemented by Gronewold et al. (2017), was used for this study. LBRM-CC-lumped is currently in use as one of the models contributing guidance to the U.S. Army Corps of Engineers Detroit District's contribution to the internationally coordinated 6-month water level forecast (Fry et al., 2020). Tab. S3 provides a list of all the LBRM-CC-lumped model parameters which were optimized for this study.

Table S3: **The parameters calibrated for the LBRM-CC-lumped model.** The table specifies the parameters calibrated for the LBRM-CC-lumped models. The "Param" column corresponded to the parameter definition with the Fortran source code, and "LBRM Params" corresponds to parameter names within the input file used for initiating model values for LBRM-CC-lumped. The parameters are uniformly distributed in the range given. The unit is specified for each parameter.

Param.	Range	Unit	LBRM Params
TBase	$[1.1, 48.0]$	1/°C	TBase
AlbedS	$[0.2, 1.0]$	cm/d/°C	Snowmelt Factor
AlpPer	$[3.0 \times 10^{-7}, 9.4]$	1/d	Linear Reservoir Coefficient (LRC) Percolation
AlpUEv	$[1.0 \times 10^{-12}, 9.9 \times 10^{-6}]$	$1/m^3$	Partial LRC Upper Soil Zone Evaporation
AlpInt	$[1.0 \times 10^{-5}, 27.0]$	1/d	LRC Interflow
AlpDpr	$[1.0 \times 10^{-5}, 93.0]$	1/d	LRC Deep Percolation
AlpLEv	$[1.0 \times 10^{-13}, 0.05]$	$1/m^3$	Partial LRC Lower Soil Zone Evaporation
AlpGw	$[1.0 \times 10^{-7}, 1.0]$	1/d	LRC groundwater
AlpSf	$[0.02, 3.0]$	1/d	LRC surface flow
USZC	$[1.8, 25.0]$	cm	Upper Soil Zone Capacity

S.2.3 HYMOD2-lumped

HYMOD is a conceptual rainfall-runoff model introduced by Boyle et al. (2000). In this study, we use the modified lumped
85 version of the HYMOD model, called HYMOD2. The model is referred to as "HYMOD2-lumped" in this study to follow the same naming convention for all models indicating whether models are lumped or distributed. The HYMOD2-lumped model is working in reliance on the probability distributed storage capacity concept proposed by Moore (1985), representing the vertical soil moisture accounting process. HYMOD is originally designed as a lumped model utilizing Nash cascade for horizontal routing and a leaky linear reservoir to represent baseflow.
90 The modified HYMOD2-lumped version used in this study contains an improved parameterization for the evaporation process as proposed by Roy et al. (2017). Since snow is an important factor driving the water cycle in the basins considered in this model intercomparison study, we also included the Degree Day Snow model (Martinec, 1975) and rain-snow-partitioning in the lumped version of HYMOD2. Potential evapotranspiration is derived from minimum and maximum temperature as well as the basin latitude using the Hargreaves-Samani method (Hargreaves and Samani, 1985).
95 The eleven parameters calibrated for this model are given in Tab. S4.

Table S4: **The parameters calibrated for the HYMOD2-lumped model.** The table specifies the eleven parameters calibrated for the HYMOD2-lumped models. The parameters are uniformly distributed in the range given. A brief description of the parameter as well as its unit is specified.

Param.	Range	Unit	Parameter description
H	$[100, 5000]$	mm	Height of soil moisture accounting tank
B	$[0.01, 2.0]$	-	Distribution function shape parameter
α	$[0.01, 1.0]$	-	Quick-slow split parameter
Nq	$[1, 6]$	-	Number of quickflow routing tanks
Ks	$[0.001, 0.95]$	mm/d	Slowflow routing tanks rate parameter
Kq	$[0.05, 0.95]$	mm/d	Quickflow routing tanks rate parameter
Kmax	$[0.0, 1.0]$	mm/d	Upper limits for resistance to ET flux K
γ	$[0, 1]$	-	Lower limits for resistance to ET flux K as Kmin = γ Kmax
BE	$[0.01, 1.95]$	-	Power coefficient
DDF	$[0.0, 10.0]$	mm/d/°C	Degree day factor (DDF) for snowmelt
Tbase	$[-2.0, 2.0]$	°C	Base temperature for melt calculation in DDF model

S.2.4   GR4J-lumped

The GR4J model (Perrin et al., 2003) is a lumped water balance model that relates runoff to rainfall and evapotranspiration using daily data while the model contains two stores and originally has four parameters (X_1 to X_4). The GR4J parameter X_1 denotes the production store capacity, parameter X_2 is the inter-catchment exchange coefficient, X_3 is the routing store capacity, and X_4 is the unit hydrograph time constant. GR4J is used in concert with CemaNeige (Valéry et al., 2014) for the handling of snow. GR4J and CemaNeige are fully emulated within the Raven modeling framework (Craig et al., 2020) and used as such in this study. The parameters, their ranges and where these parameters are located in a Raven setup is listed in Table S5. Please note that parameter x_5 for an estimate of annual average snow is usually not part of the CemaNeige model but was added here as Raven requires an estimate and none was at hand. Two additional parameters for rain-snow partitioning were added as only precipitation forcings instead of rain and snow forcings were available.

Table S5: **The parameters calibrated for the GR4J-lumped model.** The parameters are uniformly distributed in the range given. The Raven table and parameter name can be used to locate the parameter in the Raven setup files.

Param.	Range	Unit	Raven table	Parameter name
GR4J parameters:				
x_1	$[0.01, 2.5]$	m	SoilProfiles	thickness top soil layer (GR4J X_1)
x_2	$[-15.0, 10.0]$	mm/d	SoilParameterList	GR4J_X2
x_3	$[10.0, 700.0]$	mm	SoilParameterList	GR4J_X3
x_4	$[0.0, 10.0]$	d	LandUseParameterList	GR4J_X4
CemaNeige parameters:				
x_5	$[1.0, 30.0]$	mm	GlobalParameter	AvgAnnualSnow
x_6	$[0.0, 1.0]$	1/d	GlobalParameter	AirSnowCoeff
x_7	$[2.0, 8.0]$	mm/d/°C	LandUseParameterList	MELT_FACTOR
Rain-snow partitioning:				
x_8	$[-3.0, 3.0]$	°C	GlobalParameter	RAINSNOW_TEMP
x_9	$[0.5, 6.0]$	°C	GlobalParameter	RAINSNOW_DELTA

S.2.5 HMETS-lumped

HMETS is a very simple, yet efficient, lumped conceptual hydrological model simulating all main hydrological processes including snow accumulation and snowmelt (Martel et al., 2017). HMETS us setup through the Raven hydrologic modeling framework which is able to fully emulate the original HMETS implementation. The parameters calibrated for HMETS are listed in Table S6.

Table S6: **The parameters calibrated for the HMETS-lumped model.** The parameters are uniformly distributed in the range given. The Raven table and parameter name can be used to locate the parameter in the Raven setup files. A two-layer soil model was used here. The TOPSOIL is the upper soil layer while PHREATIC is the lower soil layer. The two Raven parameters, SNOW_SWI_MAX and MAX_MELT_FACTOR, are derived using a sampled parameter (x_6, x_{10}) and SNOW_SWI_MIN and MIN_MELT_FACTOR, respectively, to make sure that one parameter is always larger than the other.

Param.	Range	Unit	Raven table	Parameter name
x_1	$[0.3, 20.0]$	-	LandUseParameterList	GAMMA_SHAPE
x_2	$[0.01, 5.0]$	-	LandUseParameterList	GAMMA_SCALE
x_3	$[0.5, 13.0]$	-	LandUseParameterList	GAMMA_SHAPE2

Continued on next page

Param.	Range	Unit	Raven table	Parameter name
x_4	$[0.15, 1.5]$	-	LandUseParameterList	GAMMA_SCALE2
x_5	$[0.0, 20.0]$	mm/d/°C	LandUseParameterList	MIN_MELT_FACTOR
x_6	$[0.0, 20.0]$	mm/d/°C	LandUseParameterList	MAX_MELT_FACTOR =
				MIN_MELT_FACTOR + x_6
x_7	$[-2.0, 3.0]$	°C	LandUseParameterList	DD_MELT_TEMP
x_8	$[0.01, 0.2]$	1/mm	LandUseParameterList	DD_AGGRADATION
x_9	$[0.0, 0.1]$	frac	GlobalParameter	SNOW_SWI_MIN
x_{10}	$[0.01, 0.3]$	frac	GlobalParameter	SNOW_SWI_MAX =
				SNOW_SWI_MIN + x_{10}
x_{11}	$[0.005, 0.1]$	1/mm	GlobalParameter	SWI_REDUCT_COEFF
x_{12}	$[-5.0, 2.0]$	°C	LandUseParameterList	DD_REFREEZE_TEMP
x_{13}	$[0.0, 5.0]$	mm/d/°C	LandUseParameterList	REFREEZE_FACTOR
x_{14}	$[0.0, 1.0]$	-	LandUseParameterList	REFREEZE_EXP
x_{15}	$[0.0, 3.0]$	-	SoilParameterList	PET_CORRECTION TOPSOIL
x_{16}	$[0.0, 1.0]$	-	LandUseParameterList	HMETS_RUNOFF_COEFF
x_{17}	$[0.00001, 0.02]$	1/d	SoilParameterList	PERC_COEFF TOPSOIL
x_{18}	$[0.0, 0.1]$	1/d	SoilParameterList	BASEFLOW_COEFF TOPSOIL
x_{19}	$[0.00001, 0.01]$	1/d	SoilParameterList	BASEFLOW_COEFF PHREATIC
x_{20}	$[0.0, 0.5]$	m	SoilProfiles	thickness TOPSOIL
x_{21}	$[0.0, 2.0]$	m	SoilProfiles	thickness PHREATIC
x_{22}	$[-3.0, 3.0]$	°C	GlobalParameter	RAINSNOW_TEMP
x_{23}	$[0.5, 6.0]$	°C	GlobalParameter	RAINSNOW_DELTA

S.2.6 Blended-lumped

115 The lumped Blended Model is a hydrologic model defined within the Raven hydrologic modeling framework (Craig et al., 2020). The Blended Model uses the weighted average of several chosen process implementations for key processes to simulate streamflow instead of using one single parametrization per process.

As an example, a (simplified) blended model could be defined as:

$$f_{shared}(x, w) = (w_{d1}D_1 + w_{d2}D_2) \cdot (w_{e1}E_1 + w_{e2}E_2 + w_{e3}E_3) + (w_{f1}F_1 + w_{f2}F_2) \tag{S1}$$

120   with

$$w_{d1} + w_{d2} = 1 \tag{S2}$$

$$w_{e1} + w_{e2} + w_{e3} = 1 \tag{S3}$$

$$w_{f1} + w_{f2} = 1 \tag{S4}$$

where D_1 and D_2 could be, for instance, be two options for one process. For example, deriving infiltration could be per-
125   formed once using the infiltration definition of HMETS (D_1) and once derived as defined in the HBV model (D_2). The
infiltration outputs D_1 and D_2 are then weighted using w_{d1} and w_{d2} to derive the infiltration estimate Raven will use for the
remainder of the simulation. A detailed description of all processes and process options as well as a model flowchart can be
found in the Supplementary Material of Mai et al. (2020) and in the Raven documentation (Craig et al., 2020).

The Blended Model was introduced by Mai et al. (2020), analysed in calibration mode by Chlumsky et al. (2021), and
130   deployed across North America by Mai et al. (2022). The exact same model setup was employed here. It uses three different
options M_i for the infiltration process, three options N_i for quickflow, two options O_i for evaporation, two options P_i for
baseflow, and three options Q_i for snow balance. All other processes, i.e., convolution of surface runoff R_1 and delayed runoff
S_1, potential melt T_1 , percolation U_1 , rain–snow partitioning V_1 , and precipitation correction W_1, are used with one fixed
process option. The remaining processes also have only one option, but none of them contains tunable parameters. They are
135   merged to a "remaining" process X_1, which will never appear in the sensitivity analysis because it is constant. When the first
option of each of the processes M_1, N_1, O_1, P_1, Q_1, R_1, S_1, T_1, U_1, V_1, W_1, and X_1, is chosen and parameter x_{35} is set
to zero, the Raven setting emulates the HMETS model (Martel et al., 2017) perfectly. All other combinations are unnamed
models.

Details of process options and parameters can be found in Appendix C (Tables C1 and C2) in Mai et al. (2020). The list of
140   parameters is added to the Supplementary Material herein (Table S7) for the convenience of the readers.

Table S7: **The parameters calibrated for the Blended-lumped model.** The parameters are uniformly distributed in the range given. The Raven table and parameter name can be used to locate the parameter in the Raven setup files. A three-layer soil model was used here, with the third (groundwater) layer being of infinite depth. The TOPSOIL is the upper soil layer while PHREATIC is the lower soil layer. The three Raven parameters, FIELD_CAPACITY TOPSOIL, SNOW_SWI_MAX, and MAX_MELT_FACTOR, are derived using a sampled parameter (x_{10}, x_{14}, and x_{25}) and SAT_WILT TOPSOIL, SNOW_SWI_MIN, and MIN_MELT_FACTOR, respectively, to make sure that one parameter is always larger than the other. The baseflow coefficients, BASEFLOW_COEFF TOPSOIL and PHREATIC, are derived from parameters x_4 and x_{11} to allow a logarithmic sampling. This table is taken from Mai et al. (2020) (Appendix C Table C2 therein) and augmented by the weight generating parameters (r_1 to r_8) that are used to describe the model structure.

Param.	Range	Unit	Raven table	Parameter name
Infiltration:				
x_1	$[0.0, 1.0]$	-	LandUseParameterList	HMETS_RUNOFF_COEFF
x_2	$[0.1, 3.0]$	-	SoilParameterList	B_EXP TOPSOIL
x_3	$[0.5, 3.0]$	-	SoilParameterList	HBV_BETA TOPSOIL
Quickflow:				
x_4	$[-5.0, -1.0]$	1/d	SoilParameterList	BASEFLOW_COEFF TOPSOIL = 10.0^{x_4}
x_5	$[0.0, 100.0]$	mm/d	SoilParameterList	MAX_BASEFLOW_RATE TOPSOIL
x_6	$[0.5, 2.0]$	-	SoilParameterList	BASEFLOW_N TOPSOIL
x_7	$[5.0, 10.0]$	m	TerrainClasses	TOPMODEL_LAMBDA
Evaporation:				
x_8	$[0.0, 3.0]$	-	SoilParameterList	PET_CORRECTION TOPSOIL
x_9	$[0.0, 0.05]$	frac	SoilParameterList	SAT_WILT TOPSOIL
x_{10}	$[0.0, 0.45]$	frac	SoilParameterList	FIELD_CAPACITY TOPSOIL = SAT_WILT TOPSOIL + x_{10}
Baseflow:				
x_{11}	$[-5.0, -2.0]$	1/d	SoilParameterList	BASEFLOW_COEFF PHREATIC = $10.0^{x_{11}}$
x_{12}	$[0.5, 2.0]$	-	SoilParameterList	BASEFLOW_N PHREATIC
Snow balance:				
x_{13}	$[0.0, 0.1]$	frac	GlobalParameter	SNOW_SWI_MIN
x_{14}	$[0.01, 0.3]$	frac	GlobalParameter	SNOW_SWI_MAX = SNOW_SWI_MIN + x_{14}
x_{15}	$[0.005, 0.1]$	1/mm	GlobalParameter	SWI_REDUCT_COEFF
x_{16}	$[-5.0, 2.0]$	°C	LandUseParameterList	DD_REFREEZE_TEMP

Continued on next page

Param.	Range	Unit	Raven table	Parameter name
x_{17}	$[0.0, 1.0]$	-	LandUseParameterList	REFREEZE_EXP
x_{18}	$[0.0, 5.0]$	mm/d/°C	LandUseParameterList	REFREEZE_FACTOR
x_{19}	$[0.0, 0.4]$	frac	GlobalParameter	SNOW_SWI
Convolution (surface runoff):				
x_{20}	$[0.3, 20.0]$	-	LandUseParameterList	GAMMA_SHAPE
x_{21}	$[0.01, 5.0]$	-	LandUseParameterList	GAMMA_SCALE
Convolution (delayed runoff):				
x_{22}	$[0.5, 13.0]$	-	LandUseParameterList	GAMMA_SHAPE2
x_{23}	$[0.15, 1.5]$	-	LandUseParameterList	GAMMA_SCALE2
Potential melt:				
x_{24}	$[1.5, 3.0]$	mm/d/°C	LandUseParameterList	MIN_MELT_FACTOR
x_{25}	$[0.0, 5.0]$	mm/d/°C	LandUseParameterList	MAX_MELT_FACTOR =
				MIN_MELT_FACTOR + x_{25}
x_{26}	$[-1.0, 1.0]$	°C	LandUseParameterList	DD_MELT_TEMP
x_{27}	$[0.01, 0.2]$	1/mm	LandUseParameterList	DD_AGGRADATION
Percolation:				
x_{28}	$[0.00001, 0.02]$	1/d	SoilParameterList	PERC_COEFF TOPSOIL
x_{35}	$[0.0, 0.02]$	1/d	SoilParameterList	PERC_COEFF PHREATIC
Rain-snow partitioning:				
x_{31}	$[-3.0, 3.0]$	°C	GlobalParameter	RAINSNOW_TEMP
x_{32}	$[0.5, 4.0]$	°C	GlobalParameter	RAINSNOW_DELTA
Precipitation correction:				
x_{33}	$[0.8, 1.2]$	-	Gauge	RAINCORRECTION
x_{34}	$[0.8, 1.2]$	-	Gauge	SNOWCORRECTION
Soil model:				
x_{29}	$[0.0, 0.5]$	m	SoilProfiles	thickness TOPSOIL
x_{30}	$[0.0, 2.0]$	m	SoilProfiles	thickness PHREATIC
Model structure:				
r_1	$[0.0, 1.0]$	-	HydrologicProcesses	weight gen. parameter INFILTRATION
r_2	$[0.0, 1.0]$	-	HydrologicProcesses	weight gen. parameter INFILTRATION
r_3	$[0.0, 1.0]$	-	HydrologicProcesses	weight gen. parameter QUICKFLOW
r_4	$[0.0, 1.0]$	-	HydrologicProcesses	weight gen. parameter QUICKFLOW

Param.	Range	Unit	Raven table	Parameter name
r_5	$[0.0, 1.0]$	-	HydrologicProcesses	weight gen. parameter SOILEVAP
r_6	$[0.0, 1.0]$	-	HydrologicProcesses	weight gen. parameter BASEFLOW
r_7	$[0.0, 1.0]$	-	HydrologicProcesses	weight gen. parameter SNOWBALANCE
r_8	$[0.0, 1.0]$	-	HydrologicProcesses	weight gen. parameter SNOWBALANCE

S.2.7 Blended-Raven

The Blended-Raven model is the same as the Blended-lumped model setup using the common subbasin discretization provided through the common routing product. The parameters calibrated are the ones calibrated for the Blended-lumped model (Table S7 augmented by the routing specific parameters listed in Table S8.

Table S8: **The parameters calibrated for the Blended-Raven model.** The parameters are uniformly distributed in the range given. The Raven table and parameter name can be used to locate the parameter in the Raven setup files. Only the routing parameters are explicitly listed while all the other parameters are the same as given in the table for the Blended-lumped model to avoid duplication.

Param.	Range	Unit	Raven table	Parameter name
All parameters x_1 to x_{35} and r_1 to r_8 used for Blended-lumped (Tab. S7)				
Routing:				
x_{36}	$[-1.0, 1.0]$	-	SBGroupPropertyMultiplier	MANNINGS_N = $5.0^{x_{36}}$
x_{37}	$[0.1, 2.0]$	-	SBGroupPropertyMultiplier	RESERVOIR_CREST_WIDTH

S.2.8 VIC-Raven

In GRIP-GL, the VIC and Raven routing model are coupled to simulate rainfall-runoff processes in the Great Lakes region. The VIC Image Driver of version 5.1.0 is used, and Raven is employed as a routing module for VIC. It should be noted that the drainage from the bottom soil layer in VIC is named as "baseflow" (Gao et al., 2010); however, this is a misnomer that implies water directly enters the streams without being delayed in storage (Li et al., 2015). Thus, we use the term "recharge" instead of "baseflow" to represent this flux component in this context, which is consistent with the terminology used in Raven (Craig et al., 2020; Craig, 2022).

14

In GRIP-GL, VIC model is built at the RDRS-v2 forcing grid-cells with a resolution of 15 km by 15 km. VIC requires DEM,
155   soil, and land cover data for its parameterization. The input data for VIC are the sub-daily meteorological drivers from the
RDRS-v2 forcing data set, i.e., precipitation, air temperature, atmospheric pressure, incoming shortwave radiation, incoming
longwave radiation, vapor pressure and wind speed. The VIC-generated (quick) runoff and recharge fluxes at the grid-cell scale
are firstly aggregated into the sub-watershed scale, and then routed to the catchment outlet in terms of in-catchment routing
and river channel routing processes.

160   Raven routing module consists of in-catchment routing and channel routing processes. VIC-generated runoff is firstly simu-
lated by an in-catchment routing process, which employs a Gamma unit hydrograph to represent the time delay between runoff
and reaching the stream. It then employs channel routing using the diffusive wave method to simulate flood wave propagation
through the reach. Baseflow is simulated using a linear storage model to delay the recharge component. If catchments contain
lakes, the lake release is solved by a two-parameter lake-type reservoir model in Raven. The routing network for the Raven
165   routing module is produced by BasinMaker (Han, 2021; Han et al., 2021), and inputs for routing are subbasin-averaged runoff
and baseflow generated by VIC.

VIC has nine parameters for calibration and Raven routing module has five parameters for catchments with lakes and four
parameters for catchments without lakes for calibration (Tab. S9). These parameters are calibrated simultaneously using DDS
optimization algorithm (i.e., the VIC and Raven are coupled in calibration, and the outputs of VIC are used as inputs for
170   Raven routing). The KGE is employed as the calibration objective. The optimization is repeated for 20 trials, each with 1,000
calibration iterations, and the best results out of 20 are reported.

Table S9: **The parameters calibrated for the VIC-Raven model.** The parameters are uniformly distributed in the range given.
The unit and a brief description of each parameter is given. The parameters are grouped into the nine parameters of the VIC
land-surface scheme and the five parameters calibrated for the Raven routing.

Param.	Range	Unit	Parameter description
VIC Land-Surface Scheme:			
infilt	$(0, 5]$	-	Parameter used to describe Variable Infiltration Curve; higher values will produce more runoff
Ds	$(0, 1]$	-	Fraction of Dsmax parameter at which non-linear recharge occurs
Dsmax	$(0, 30]$	mm/d	Maximum velocity of recharge for each grid cell
Ws	$(0, 1]$	-	Fraction of maximum soil moisture where non-linear recharge occurs
d1	$[0.01, 0.5]$	m	Depth of top layer
d2	$[0.01, 2.0]$	m	Depth of first layer
d3	$[0.5, 3.0]$	m	Depth of second layer
expt	$[3, 30]$	-	Exponent parameter of Brooks-Corey relationship
Ksat	$[100, 1200]$	mm/d	Saturated hydraulic conductivity for each layer

Continued on next page

Param.	Range	Unit	Parameter description
Raven Routing:			
Mcoeff	[0.2, 5.0]	-	Multiplier to calibrate Manning's n
Bcoeff	[0.1, 0.999]	1/d	Baseflow coefficient k
Shape	[1.1, 11.0]	-	Shape parameter of the gamma unit hydrograph
Tpeak	[0.2, 2.0]	d	Time to peak parameter; used to estimate scale parameter of gamma unit hydrograph
Cwidth	[0.5, 1.5]	-	Multiplier to calibrate crest width in the lake-like reservoir model; parameter ignored if catchment has no lakes

S.2.9   SWAT-Raven

SWAT simulated vertical flux (total runoff) were fed to the lake and river routing product (Han et al., 2020), integrated into the
175   Raven modelling framework (Craig et al., 2020), to route the both the vertical flux. The integrated SWAT-Raven model was then calibrated by calibrating SWAT-related and Raven-related parameters simultaneously within the given range as shown in Tab. S10.

Table S10: **The parameters calibrated for the SWAT-Raven model.** The parameters are uniformly distributed in the range given. The unit and a brief description of each parameter is given. The parameters are grouped into the 17 parameters of the SWAT model and the two parameters calibrated for the Raven routing.

Param.	Range	Unit	Parameter description
SWAT related:			
SFTMP	[−5, 5]	°C	Snow fall temperature
SMTMP	[−5, 5]	°C	Snowfall melt base temperature
SMFMX	[2, 10]	mm/°C/d	Maximum melt rate for snow during the year
SMFMN	[2, 10]	mm/°C/d	Minimum melt rate for snow during the year
TIMP	[0, 1]	-	Snow pack temperature lag factor
CN2	[−20, 20]	%	Percent change of initial SCS runoff curve number for moisture condition II [-]
GW_DELAY	[1, 180]	d	Groundwater delay time
ALPHA_BF	[0.01, 1]	-	Base flow alpha factor

Param.	Range	Unit	Parameter description
GWQMN	[10, 1000]	mm	Threshold depth of water in the shallow aquifer required for return flow to occur
GW_REVAP	[0.02, 0.2]	-	Groundwater revap. coefficient
REVAPMN	[10, 1000]	mm	Threshold depth of water in the shallow aquifer for 'revap' to occur
SURLAG	[0.1, 24]	d	Surface runoff lag time
ESCO	[0.5, 1]	-	Soil evaporation compensation factor
EPCO	[0, 1]	-	Plant uptake compensation factor
OV_N	[0.01, 0.25]	-	Manning's n value for overland flow
SOL_K	[−20, 20]	%	Percent change of initial soil hydraulic conductivity [mm/d]
SOL_AWC	[−20, 20]	%	Percent change of initial soil available water storage capacity [-]
Raven Routing:			
w	[0.25, 4.0]	-	Multiplier to adjust crest width of reservoir/lake
n	[0.25, 4.0]	-	Multiplier to adjust Manning's coefficient

S.2.10 WATFLOOD-Raven

180 WATFLOOD simulated gridded vertical fluxes (runoff and recharge to lower zone storage) were fed to the lake and river routing product (Han et al., 2020), integrated into the Raven modelling framework (Craig et al., 2020), to route the both vertical fluxes. The integrated WATFLOOD-Raven model was then calibrated by using WATFLOOD-related and Raven-related parameters simultaneously within the given range as shown in Tab. S11.

Table S11: **The parameters calibrated for the WATFLOOD-Raven model.** The parameters are uniformly distributed in the range given. The unit and a brief description of each parameter is given. The parameters are grouped into the eleven parameters of the WATFLOOD model and the six parameters calibrated for the Raven routing.

Param.	Range	Unit	Parameter description
WATFLOOD related:			
fm	[0.05, 0.5]	mm/°C/h	Melt factor
base	[−5, 5]	°C	Base temperature
rec	[0.01, 10]	-	Interflow coefficient
ak	[0.01, 10]	-	Infiltration coefficient in bare ground

Continued on next page

Param.	Range	Unit	Parameter description
akfs	$[0.001, 1]$	-	Infiltration coefficient in snow covered ground
retn	$[50, 1000]$	mm	Upper zone retention
ak2	$[0.01, 1]$	-	Recharge coefficient in bare ground
ak2fs	$[0.001, 0.1]$	-	Recharge coefficient in snow covered ground
r3	$[1, 50]$	-	Overland flow roughness coefficient in bare ground
r3fs	$[1, 10]$	-	Overland flow roughness coefficient in snow covered ground
fpet	$[1, 20]$	-	Interception evaporation factor
Raven Routing:			
w	$[0.25, 4.0]$	-	Multiplier to adjust crest width of reservoir/lake
n	$[0.25, 4.0]$	-	Multiplier to adjust Manning's coefficient
γ-shape	$[1, 10]$	-	Gamma unit hydrograph shape parameter
γ-scale	$[1, 10]$	-	Gamma unit hydrograph scale parameter
bf_k	$[10^{-5}, 1]$	1/d	Baseflow coefficient
bf_n	$[1, 5]$	1/d	User specified soil parameter

S.2.11   MESH-CLASS-Raven

The integrated MESH-CLASS-Raven model was then calibrated by using MESH-CLASS-related and Raven-related parameters simultaneously within the given range as shown in Tab. S12.

Table S12: **The parameters calibrated for the MESH-CLASS-Raven model.** The parameters are uniformly distributed in the range given. The unit and a brief description of each parameter is given. The parameters are grouped into the 14 parameters of the MESH-CLASS model and the 6 parameters calibrated for the Raven routing.

Param.	Range	Unit	Parameter description
MESH-CLASS related (all parameters per GRU):			
DRN	$[0.001, 5.0]$	-	Soil drainage index
DDEN	$[1.0, 150.0]$	-	Drainage density
KSAT	$[1 \times 10^{-11}, 1 \times 10^{-2}]$	m/s	Saturated hydraulic conductivity
QA50	$[1.0, 100.0]$	W/m^2	Reference value of incoming shortwave radiation
ROOT	$[0.001, 5.0]$	m	Annual maximum rooting depth of vegetation category

Param.	Range	Unit	Parameter description
RSMN	$[10, 450]$	s/m	Minimum stomatal resistance of vegetation category
SDEP	$[0.001, 4.0]$	m	Soil permeable depth
WFR2	$[0.01, 2.0]$	-	Channel roughness coefficient
SAND	$[0.0, 100.0]$	%	Percentage sand content
CLAY	$[0.0, 100.0]$	%	Percentage clay content
XSLP	$[0.0001, 1.0]$	%	Average overland slope
ZSLN	$[0.01, 1.0]$	m	Limiting snow depth below which coverage is less than 100%
ZPLS	$[0.01, 1.0]$	-	Maximum water ponding depth for snow-covered areas
ZPLG	$[0.01, 1.0]$	-	Maximum water ponding depth for snow-free areas
Raven Routing:			
w	$[0.1, 5.0]$	-	Multiplier to adjust crest width of reservoir/lake
n	$[0.1, 5.0]$	-	Multiplier to adjust Manning's coefficient
γ-shape	$[1.0, 10.0]$	-	Gamma unit hydrograph shape parameter
γ-scale	$[1.0, 10.0]$	-	Gamma unit hydrograph scale parameter
bf_k	$[1 \times 10^{-5}, 1.0]$	1/d	Baseflow coefficient
bf_n	$[1.0, 5.0]$	1/d	User specified soil parameter

S.2.12 MESH-SVS-Raven

190 The MESH-SVS-Raven model was calibrated by using MESH-SVS-related and Raven-related parameters simultaneously within the given range as shown in Tab. S13. Most of the calibrated parameters correspond to coefficients that are used to multiply the actual corresponding Model parameter values. This is done in order to limit the number of free parameters during calibration, while being able to change the actual parameter values over the full domain and maintain a spatial variability and coherence for the actual field of parameter values across the domain. This approach has already proven effective in the sense that

195 model performances could be significantly improved compared to the default version of the model, that the issue of parameter value equifinality was limited, and that model robustness was satisfying. See for example the studies of Gaborit et al. (2015), Gaborit et al. (2017), and Mai et al. (2021), where some models were calibrated based on multiplying coefficients instead of actual parameter values. One other advantage of such an approach is that none of the MESH-SVS-Raven free parameters used in calibration are tied to a specific calibration or validation basin, but are applied to the whole region at once. Therefore, no

200 spatial parameter transfer is required for the validation basins, as long as they are included in the region of interest. Among the MESH-SVS-Raven parameters, only the "RICK", "FLZCOEFF", "PWRC", "GASH" and "GASC" parameters (see Tab. S13)

correspond to actual model parameter values, but here they correspond in the model to values that are fixed over the whole region where the model is applied.

In order to use different parameter values for the agricultural and more natural areas (like grassland, forests, etc.), two different multipliers were used both for the horizontal and vertical model conductivities, as well as for the evapotranspiration resistance multipliers (see Tab. S13). This was done because of the artificial tile drains that are often installed in agricultural areas of the Great-Lakes basin (see Valayamkunnath et al., 2020). These tile drains allow to prevent the crops from being flooded by quickly removing water that infiltrated into the soil column, but lead to significantly higher peak flows at the outlet of agricultural basins, in comparison to their natural regime. Because these tile drains are not explicitly represented in MESH-SVS-Raven, their effect on simulated flows was represented by allowing hydraulic conductivity values to be significantly increased in the model, as suggested by De Schepper et al. (2015). This is why higher ranges were allowed for the hydraulic conductivity multiplier values associated with agricultural areas, compared to the natural areas (see Tab. S13). However, since MESH-SVS-Raven cannot simulate processes separately for these two types of areas inside a grid-cell (SVS soil parameters are valid over the whole grid cell), the actual multiplier value used to adjust SVS horizontal or vertical hydraulic conductivity consists of a weighted average of the agricultural and "natural" multiplier values, based on the fraction of the grid-cell covered with agricultural cover, and the rest of the pixel that is assumed as being "natural", in the sense that it is assumed not to contain any tile drains. The same was done for the evapotranspiration resistance multipliers, in order to give SVS some flexibility by using different adjustments in agricultural and other (natural) areas, for these parameters.

Finally, because we rely here on multiplicative coefficients to adjust the actual parameter values, some restrictions had to be imposed on the final (after adjustment) actual model parameter values, to ensure that they could not go beyond values with physical meaning. For example, the Albedo final values have a maximum of 1, the 50% root depth values have to stay below the 95% root depth values (minus 5cm to ensure numerical stability), and the three soil moisture thresholds, namely the wilting point, the field capacity and the saturation were respectively constrained below fractions of 0.5, 0.75 and 0.95. Because they were adjusted all together by the same multiplier, their relative order remained unchanged after adjustment and did not need to be constrained.

Table S13: **The parameters calibrated for the MESH-SVS-Raven model.** The parameters are uniformly distributed in the range given. The unit and a brief description of each parameter is given. The parameters are grouped into the 17 parameters of the MESH-SVS model and the 6 parameters calibrated for the Raven routing. Most of these parameters correspond to multiplying coefficients. See text for more details.

Param.	Range	Unit	Parameter description
MESH-SVS related:			
MLTM	$[0.5, 4]$	-	Snowmelt rate divider
BMOD	$[0.5, 4]$	-	Slope of retention curve multiplier
WMOD	$[0.1, 3]$	-	Soil moisture thresholds multiplier

Continued on next page

20

Param.	Range	Unit	Parameter description
PMOD	$[0.1, 5]$	-	Soil water suction near saturation multiplier
GRKMO_N	$[0.2, 5]$	-	Horizontal hydraulic conductivity multiplier (natural areas)
GRKMO_A	$[0.2, 40]$	-	Horizontal hydraulic conductivity multiplier (agricultural areas)
KASMO_N	$[0.2, 5]$	-	Vertical hydraulic conductivity multiplier (natural areas)
KASMO_A	$[0.1, 40]$	-	Vertical hydraulic conductivity multiplier (agricultural areas)
RICK	$[0, 1]$	-	Weight given to new evaporation formulation over bare ground
EVM_N	$[0.2, 5]$	-	Evapo-transpiration resistance multiplier (natural areas)
EVM_A	$[0.2, 9]$	-	Evapo-transpiration resistance multiplier (agricultural areas)
SUMOD	$[0.1, 5]$	-	Sublimation resistance multiplier
RTMOD	$[0.1, 3]$	-	95% root depth multiplier
DMOD	$[0.1, 4]$	-	50% root depth multiplier
LMOD	$[0.2, 4]$	-	Leaf-Area Index multiplier
AMOD	$[0.1, 5]$	-	Albedo multiplier
ZMOD	$[0.2, 5]$	-	Surface Roughness multiplier
Raven Routing:			
FLZCOEFF	$[1.0 \times 10^{-8}, 1.0 \times 10^{-3}]$	-	Baseflow equation multiplicative parameter
PWRC	$[1.5, 7]$	-	Baseflow equation power parameter
R1NC	$[0.1, 9]$	-	Manning values multiplier
GASH	$[1, 32]$	-	Gamma Unit Hydrograph shape parameter
GASC	$[1, 6]$	-	Gamma Unit Hydrograph scale parameter
LACRWD	$[1.0 \times 10^{-3}, 1.0 \times 10^3]$	-	Lakes' outlet width multiplier

S.2.13 GEM-Hydro-Watroute

The GEM-Hydro-Watroute model was not calibrated directly, because it is too expensive in terms of computation time. For example, a 10-year open-loop simulation with GEM-Hydro-Watroute over the whole Great-Lakes watersheds is about a 100 times slower than a MESH-SVS-Raven similar run. This is due to several reasons. Despite the surface component of GEM-Hydro-Watroute (named GEM-Surf) is heavily parallelized (with MPI, where a domain grid can be split into numerous subgrids during computation) and can for example perform simulations over very large domains (even global simulations) in an efficient manner, it currently relies on 24-h integration cycles, leading to a lot of time being spent in Input/Output (I/O) processes, and in model jobs being submitted for each day of simulation on ECCC supercomputers, with the job sometimes waiting in a queue. In contrast, the surface component of MESH-SVS-Raven (namely MESH-SVS), which is also parallelized with an

MPI approach, is run with a single model integration over the full period, such that less time is therefore spent in I/O and job submission processes. Therefore, a MESH simulation is actually much faster than a GEM-Surf one, when running over grids with a limited number of points, like this is the case for the six regions of the Great-Lakes domain used here, with a 10 km resolution. In this study, the MESH-SVS simulation was about 40 times faster than the GEM-Surf simulation, over one

240   of the six regions and the 10 km resolution considered here, despite both models were run here on the same CPUs, with 10 CPUs for MESH-SVS. Using 4 CPUs for GEM-Surf was close to the optimum in this case: if asking for more CPUs, the job could wait more time in the queue during submission. For MESH-SVS, computation time may be further improved with using more CPUs, but probably not to a significant degree, since the six regions used here do not contain a high number of 10 km grid points (the maximum is 4455 grid-points for the Lake Superior region, while the minimum is 2132 for the Ottawa River

245   region). Therefore, it is very challenging to calibrate GEM-Hydro-Watroute directly already because of its surface component.

Regarding the routing component of GEM-Hydro-Watroute (Watroute), it is also much slower than the Raven routing scheme. First, Watroute is a grid-based model that has a 1 km resolution in GEM-Hydro-Watroute, and that is not parallelized. In the ECCC National Surface and River Prediction System (NSRPS; Durnford et al., 2021) that relies on the GEM-Hydro-Watroute model, the routing Watroute component is however implemented separately for each of six main Canadian Watersheds

250   that it currently covers in the NSRPS system, leading to a parallelization of the model by separate model setups. In opposition, Raven is a subbasin-based routing model. Therefore, it generally includes much less subbasin objects than there are 1 km grid points inside a given watershed. Raven to Watroute comparisons performed as part of collaborative University of Waterloo-ECCC research projects showed that Raven is about 30 times faster than Watroute, over a 60,000 km^2 watershed, with a 6-h Raven time-step and the detailed Raven version, which had 2635 objects compared to Watroute's 122,441 grid points included

255   in the basin. But this ratio is based on the model computing time only. However, running Watroute inside of GEM-Hydro-Watroute also implies a pre-processing phase of GEM-Surf outputs, which can be even more time-consuming than the model runtime itself. When including Watroute's pre-processing time, Raven was about 130 times faster than Watroute, for the basin described above. However, when comparing Watroute to Raven runtimes on the Lake Huron region of this GRIP-GL project, Raven was able to perform the 10-y calibration period in 130s, while it took about 55h to do the same for Watroute, which

260   represents a ratio of about 1500. It is true that in this case, Watroute was running over the full Lake Huron watershed, while Raven was only simulating the calibration and validation basins, which in total, represent about half of the Lake Huron region. Moreover, during these 55h of the total Watroute runtime, about 76% was spent on the pre-processing, highlighting where the emphasis should be put in order to decrease Watroute computation time. It was noticed that Watroute pre-processing time could be improved significantly, following some previous tests. But even if considering only Watroute true model computation

265   time over half of Lake Huron region, Raven would still be about 200 times faster than Watroute in this case. The difference with the ratio of 30 mentioned above probably comes from the fact that the common Raven setup used for GRIP-GL does not correspond to the most possible detailed version of the model, because lakes smaller than 5 km^2 were removed from the setup, allowing to significantly reduce the total number of subbasins in the Raven setup used for GRIP-GL. Nevertheless, the Raven setup used here proved to be able to simulate hydrographs at gauge locations in a very satisfactory manner during this project.

270   However, as mentioned in the main part of this study, GEM-Hydro-Watroute computation time is generally not an issue when it comes to perform large-scale forecasts, where the lead-time is generally limited to about 16 days. GEM-Hydro-Watroute can still perform 10 years open-loop runs over large watersheds in a reasonable time (about 7 to 10 days), given that this type of simulations is generally not needed frequently and that tests needed for model development can be done over shorter periods and smaller domains. It has also to be mentioned that Watroute pre-processing time is not linearly correlated with the size

275   of the domain, and large domains can have a pre-processing time quite similar to the one of small domains. However, this computation time is definitely an issue when it comes to model calibration. Despite different approaches have been tried in the past to calibrate GEM-Hydro-Watroute (see main part of the manuscript), in this study the idea is to emulate GEM-Surf with MESH-SVS, and to replace Watroute by Raven during calibration. This is done because using a detailed routing model is still wanted in order to be able to compute streamflow performances even for small watersheds. Moreover, it is assumed that most

280   of the bias and dynamics of streamflow simulations are controlled by the surface component, and not by the routing model, despite the latter can of course control the timing of peak flows and the dynamics of flow recessions, through the Manning and baseflow parameters, for example.

Therefore, as part of this work, the SVS-related parameters that were calibrated with MESH-SVS-Raven were transferred directly to SVS in GEM-Hydro-Watroute. For the routing parameters, only the calibrated baseflow parameters were directly

285   transferred to GEM-Hydro-Watroute afterwards. Indeed, Watroute does not rely on the Gamma Unit Hydrograph convolution to route surface runoff to the streams, and does not rely on the same lake outflow equation than Raven (thus, not all lakes are explicitly represented in Watroute), and the default Manning values in Watroute are not computed the same way as in Raven. However, because Watroute default Manning values may not be optimal, especially in conjunction with the calibrated baseflow parameters, they were manually tuned afterwards over each of the six regions, by using multiplicative coefficients applied to the

290   default (spatially varying) values (see Tab. S14), similarly to what was done during the Raven calibration. For this, a few trials (about 4 to 6 generally) were realized by running Watroute over 2 years of the calibration period, when forced by the calibrated GEM-Surf version and when using the calibrated baseflow parameters. To do so, Watroute's pre-processing therefore only had to be done once, and a two-year Watroute run without the pre-processing only took about 2.5h. The final multiplicative coefficients retained to adjust Watroute's original (default) Manning values, after these manual adjustments, range between

295   0.5 and 2 across the six regions, with the value of 1 (original Watroute Manning values remain unchanged) being retained for three out of the six regions. A limited sensitivity of the simulated flow performances to the adjusted Manning values moreover confirmed the stronger importance of the LSS parameters compared to the routing scheme (especially Manning values) parameters used here, which only control the timing of simulated streamflow to some degree, and not its general bias.

The approach employed here to calibrate GEM-Hydro-Watroute by emulating the surface component of GEM-Surf with

300   MESH-SVS and employing the faster routing scheme Raven in place of Watroute during calibration seemed relevant, as long as the two models could produce similar streamflow simulations. For this reason and during this project, all subgrid-scale lakes were neglected when simulating the surface with MESH-SVS or GEM-Surf. Indeed, with GEM-Surf (see Bernier et al., 2011), a total of five different types of surfaces can be simulated, including land, water, and urban areas, which are the three different types of surfaces present in the Great-Lakes region. However, with the current version of MESH, only the land type

305 can be simulated when using the SVS LSS. Work is under way to integrate additional GEM-Surf routines in MESH, in order to represent the other types of surfaces that GEM-Surf can simulate. Note that in the current GEM-Hydro-Watroute model used for hydrology simulations at ECCC, urban areas are not yet represented with the urban component of GEM-Surf, but are represented with SVS directly, assuming a fixed ratio of 33% imperviousness for urban cover, based on Gaborit et al. (2013). But because of this, and in order to obtain GEM-Surf and MESH-SVS simulations that are as close as possible to favor parameter transfer between the two, the subgrid-scale lakes were neglected in GEM-Surf, which means that any grid-cell having a fraction of water lower than 100% was in fact assumed to contain 100% of land instead. This resulted in almost 100% of the calibration/validation basins studied here to be covered at 100% by land in both MESH-SVS and GEM-Surf. In terms of streamflow impact, neglecting these subgrid-scale lakes in GEM-Surf only has a minor effect on GEM-Hydro-Watroute simulations across the Great Lakes, probably because of the relatively small role that lake evaporation plays in the overall

315 water balance of the terrestrial watersheds of the Great Lakes (i.e., with the exception of the five Great Lakes themselves). Moreover, since land evapotranspiration rates are generally similar to overlake evaporation in this region, replacing water surfaces with land surfaces does not have a strong impact on the resulting streamflow simulations in this region (not shown here).

However, strong differences were noticed for some flow gauges, between the simulations obtained with the default versions

320 of MESH-SVS-Raven and GEM-Hydro-Watroute. For some gauges, these differences were partly due to the fact that the Raven basin delineation was different than one used for Watroute. However, strong delineation differences between the two models was quite rare and only affected a few of the total 212 calibration/validation basins used in this project. Of course, differences can arise from the different routing models themselves, because they might not represent the same river and lake network, and because they might not represent the same processes: for example, Raven uses a Gamma Unit Hydrograph convolution

325 approach in order to route the surface runoff generated by the LSS, into the closest stream. In Watroute, surface runoff is assumed to enter the stream of the current grid-cell instantaneously, and the fact that Watroute assumes that a stream is present in every grid-cell can, to some degree, represent an approximation of the transport of surface/subsurface runoff between its source and an actual stream. However, these routing model differences should only affect streamflow timing, and not the general streamflow bias, for basins having close delineation boundaries between the two routing schemes. This affirmation

330 is especially true given that subgrid-scale lakes were neglected in the MESH-SVS and GEM-Surf surface simulations of this project (see above). Therefore, evaporation losses from lakes and rivers were not taken into account in both the Raven and Watroute simulations performed here, because such losses, when coupling Watroute or Raven to MESH-SVS or GEM-Surf, are included in the aggregated surface runoff of a surface grid cell. Surface runoff over land can however not be negative with SVS, because the evapotranspiration losses are already removed from the SVS vegetation and soil storage compartments. And

335 the main difference that can arise between Raven and Watroute simulations, when forced by the same surface fluxes and in terms of general flow bias, is due to the treatment of evaporation in the routing scheme. Therefore, in this case, because no evaporation loss is provided to either routing scheme, and because they both close their water balance almost perfectly, the routing schemes can not be responsible for strong differences regarding their flow simulations' biases (again, assuming that they have similar delineations for a given watershed, which is generally the case here).

340   Regarding MESH-SVS and GEM-Surf simulations, differences existing between the two are well known and were already
noticed during the GRIP-E project (see Mai et al., 2021). However, these differences were generally small and both models
still produced very similar flow simulations in this case. After investigation, it was found that there was an issue regarding the
reading of vegetation cover fractions from the geophysical file, in MESH-SVS-Raven. As a consequence, MESH-SVS did not
use the correct vegetation cover during GRIP-GL, and seems to have assigned the same vegetation fractions for all points inside
345   a region, based on one of the points of the region. This issue did not occur during GRIP-E, because it is related to the new
NetCDF format used in GRIP-GL for the MESH geophysical input file. This is a strong problem for the approach used here,
because SVS vegetation classes are associated to specific vegetation parameters. Therefore, the calibrated parameters related to
vegetation properties (like LAI, stomatal resistance, root dephts, etc; see Tab. S13) that were obtained with MESH-SVS-Raven
were not optimal for GEM-Hydro-Watroute.

350   This probably explains the strong flow bias differences between the default versions of the two models, for some regions, as
well as most of the strong loss of performance when transferring the SVS and some routing parameters that were calibrated
with MESH-SVS-Raven, into GEM-Hydro-Watroute. But despite this issue, the approach employed here to calibrate GEM-
Hydro-Watroute seems promising, in view of the results obtained in this study. The fact that the parameters calibrated with
MESH-SVS-Raven, when transferred into GEM-Hydro-Watroute, could still generally improve upon the default version of the
355   model, is probably due to the fact the calibrated SVS parameters related to soil processes, for example hydraulic conductivity,
were more appropriate than the default values used in GEM-Hydro, especially for agricultural areas with tile drains, that are
not represented explicitly in the model, yet. However, work is under way to fix the issue mentioned above, as well as revising
the strategy for some calibration parameters, and restart all MESH-SVS-Raven calibrations, GEM-Hydro simulations, and the
auxiliary variables' evaluation, in order to come up with a calibration strategy for GEM-Hydro-Watroute, that would allow
360   to improve or maintain the performances regarding surface fluxes and auxiliary variables, in view of being able to couple a
GEM-Hydro-Watroute model that was calibrated by targeting streamflow performances, with an atmospheric model, which is
the long-term goal of this work.

Table S14: **The parameters manually tuned for the GEM-Hydro-Watroute model.** Some of the parameters calibrated with
MESH-SVS-Raven were transferred into GEM-Hydro-Watroute, and the Watroute Manning values were then further manually
tuned afterwards. See text for more details.

Param.	Range	Unit	Parameter description
Watroute Routing:			
MULN	$[0.2, 5]$	-	Manning values multiplier

S.3 Model performance for streamflow in KGE components

365 The following figures show the model performance for the three Kling-Gupta efficiency (KGE) components of variability (KGE_α; Fig. S2), bias (KGE_β; Fig. S3) and correlation (KGE_r; Fig. S4). The components are transformed in a way that all have their optimal (maximal) value at 1 similar to the overall KGE for better comparability. The metrics are defined in the main manuscript (Eq. 1 to 3). All the component performance values can be visualized on the project website (www.hydrohub.org/grip-gl/maps_streamflow.html).

S.4 Streamflow gauging stations

370

Table S15 contains the list of the 212 gauging stations including their area, region they are located in, objectives they were selected for, and whether they were used for calibration or validation.

Figure S2. Model performance (variability) regarding streamflow. The performance is shown for (A) the 141 calibration stations and (B) the 71 validation locations for the calibration period. The results for the validation period of the calibration and validation sites are shown in panels (C) and (D), respectively. In summary, panel (B) shows spatial validation, panel (C) shows temporal validation, and panel (D) shows spatio-temporal validation across the locations (x-axis) and the 13 models (y-axis). The locations are grouped according to their location within the six watersheds (vertical black lines) of Lake Erie (ERI), Lake Huron (HUR), Lake Michigan (MIC), Lake Ontario (ONT), Ottawa River (OTT), and Lake Superior (SUP). The horizontal black lines separate the machine-learning based global LSTM model from the models that are calibrated locally and the models that are calibrated per region. The performance is quantified using the component α of the Kling-Gupta efficiency (KGE$_\alpha$) which determines the relative variability of the simulations and observations. The median KGE$_\alpha$ performance of each model for each of the four evaluation scenarios is added as labels to each panel. The thresholds for medium, good, and excellent performance classifications are added as labels to the colorbar. The performance regarding KGE can be found in Fig. 3 in the main manuscript. For the spatial distribution of these results as well as the simulated and observed hydrographs please refer to the website (www.hydrohub.org/grip-gl/maps_streamflow.html).

Figure S3. Model performance (bias) regarding streamflow. The performance is shown for (A) the 141 calibration stations and (B) the 71 validation locations for the calibration period. The results for the validation period of the calibration and validation sites are shown in panels (C) and (D), respectively. In summary, panel (B) shows spatial validation, panel (C) shows temporal validation, and panel (D) shows spatio-temporal validation across the locations (x-axis) and the 13 models (y-axis). The locations are grouped according to their location within the six watersheds (vertical black lines) of Lake Erie (ERI), Lake Huron (HUR), Lake Michigan (MIC), Lake Ontario (ONT), Ottawa River (OTT), and Lake Superior (SUP). The horizontal black lines separate the machine-learning based global LSTM model from the models that are calibrated locally and the models that are calibrated per region. The performance is quantified using the component β of the Kling-Gupta efficiency (KGE$_\beta$) which determines the bias between simulations and observations. The median KGE$_\beta$ performance of each model for each of the four evaluation scenarios is added as labels to each panel. The thresholds for medium, good, and excellent performance classifications are added as labels to the colorbar. The performance regarding KGE can be found in Fig. 3 in the main manuscript. For the spatial distribution of these results as well as the simulated and observed hydrographs please refer to the website (www.hydrohub.org/grip-gl/maps_streamflow. html).
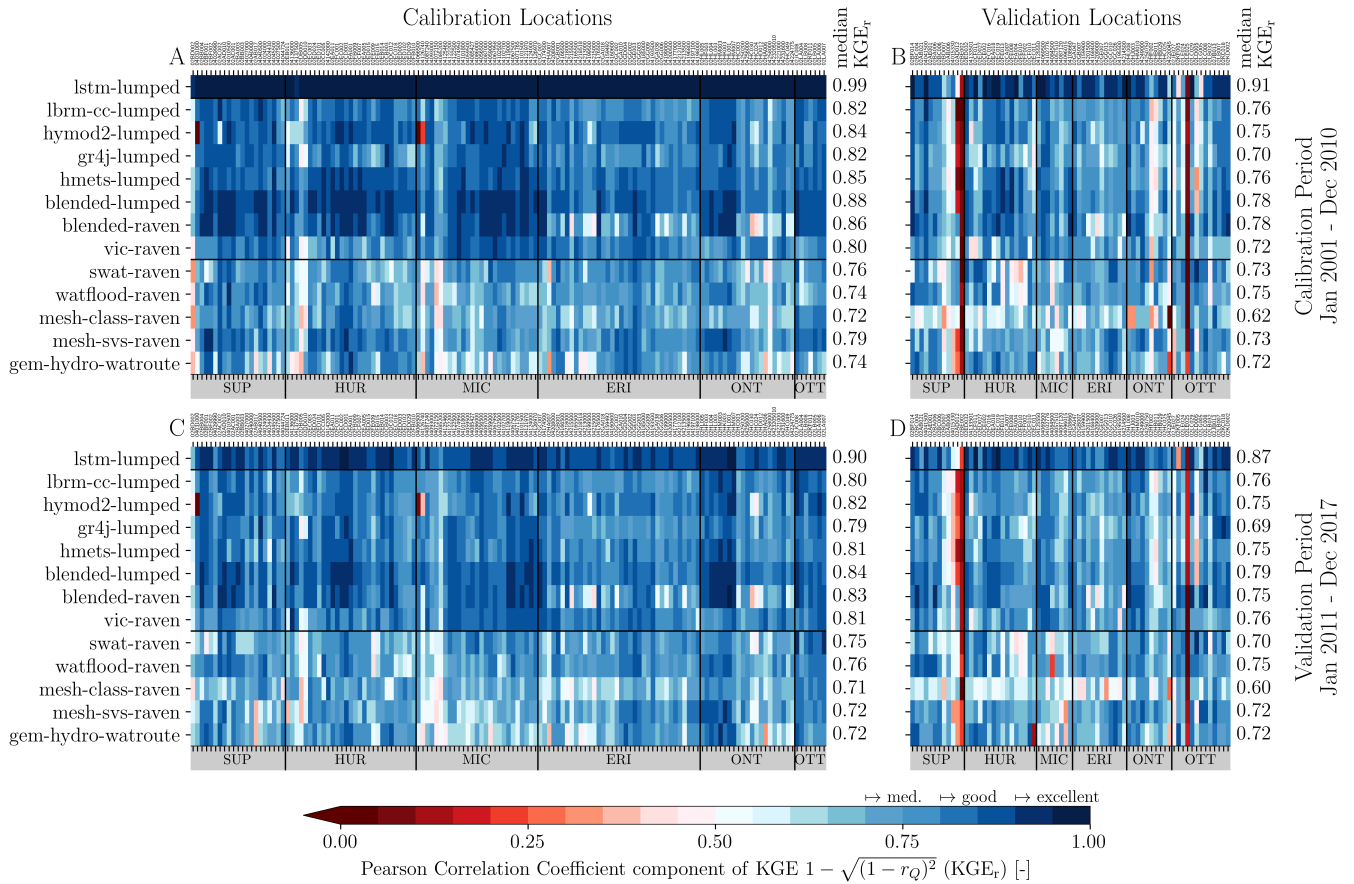
Figure S4. Model performance (correlation) regarding streamflow. The performance is shown for (A) the 141 calibration stations and (B) the 71 validation locations for the calibration period. The results for the validation period of the calibration and validation sites are shown in panels (C) and (D), respectively. In summary, panel (B) shows spatial validation, panel (C) shows temporal validation, and panel (D) shows spatio-temporal validation across the locations (x-axis) and the 13 models (y-axis). The locations are grouped according to their location within the six watersheds (vertical black lines) of Lake Erie (ERI), Lake Huron (HUR), Lake Michigan (MIC), Lake Ontario (ONT), Ottawa River (OTT), and Lake Superior (SUP). The horizontal black lines separate the machine-learning based global LSTM model from the models that are calibrated locally and the models that are calibrated per region. The performance is quantified using the component r of the Kling-Gupta efficiency (KGE$_r$) which determines the Pearson correlation between the simulations and observations. The median KGE$_r$ performance of each model for each of the four evaluation scenarios is added as labels to each panel. The thresholds for medium, good, and excellent performance classifications are added as labels to the colorbar. The performance regarding KGE can be found in Fig. 3 in the main manuscript. For the spatial distribution of these results as well as the simulated and observed hydrographs please refer to the website (www.hydrohub.org/grip-gl/maps_streamflow.html).

Table S15: **Streamflow gauges used in this study.** The 212 streamflow gauges are given including their gauge ID either specified by Water Survey Canada or U.S. Geological Survey and their name. The drainage area upstream of each gauge is specified as derived by the routing product. The region specifies in which of the six watersheds comprising the great Lakes the watershed is located in (ERI=Lake Erie, HUR=Lake Huron, MIC=Lake Michigan, SUP=Lake Superior, ONT=Lake Ontario, OTT=Ottawa River). The objective for each basin is assigned to be 1 if the watershed is of low-human impact while it is assigned to 2 if the gauge station is most downstream to one of the five lakes or the Ottawa River. Calibration basins (141) are specified with 'Cal' and validation basins (71) are specified with 'Val' in the column 'Cal/Val'. Furthermore, the mean elevation and average annual runoff of each basin is given.

Gauge ID	Name	Derived Area [km^2]	Region	Obj.	Cal/ Val	Elev. [m]	Avg. Annual Runoff [mm]
02AB006	KAMINISTIQUIA RIVER AT KAMINISTIQUIA	6576	SUP	2	Val	299	282
02AB017	WHITEFISH RIVER AT NOLALU	229	SUP	1,2	Cal	321	295
02AB021	CURRENT RIVER AT STEPSTONE	404	SUP	1,2	Val	333	335
02AC001	WOLF RIVER AT HIGHWAY NO. 17	634	SUP	1,2	Cal	184	305
02AC002	BLACK STURGEON RIVER AT HIGHWAY NO. 17	2733	SUP	1,2	Cal	188	268
02AD010	BLACKWATER RIVER AT BEARDMORE	718	SUP	1	Cal	288	353
02AD012	NIPIGON RIVER AT ALEXANDER GS	25112	SUP	2	Val	165	515
02AE001	GRAVEL RIVER NEAR CAVERS	624	SUP	1,2	Val	240	392
02BA003	LITTLE PIC RIVER NEAR COLDWELL	1378	SUP	1,2	Cal	195	364
02BA006	STEEL RIVER BELOW SANTOY LAKE	1219	SUP	1,2	Val	196	384
02BB003	PIC RIVER NEAR MARATHON	4214	SUP	1,2	Cal	189	365
02BB004	CEDAR CREEK NEAR HEMLO	214	SUP	2	Cal	297	319
02BC004	WHITE RIVER BELOW WHITE LAKE	4129	SUP	2	Val	306	352
02BC006	Pukaskwa River BELOW FOX RIVER	454	SUP	2	Val	348	472
02BD002	MICHIPICOTEN RIVER AT HIGH FALLS	5104	SUP	2	Cal	193	381
02BD007	MAGPIE RIVER NEAR WAWA	1979	SUP	2	Val	262	401
02BE002	MONTREAL RIVER NEAR MONTREAL RIVER HARBOUR	3369	SUP	2	Val	256	392
02BF001	BATCHAWANA RIVER NEAR BATCHAWANA	1225	SUP	1,2	Cal	220	541
02BF002	GOULAIS RIVER NEAR SEARCHMONT	1181	SUP	1	Cal	298	477
02BF014	GOULAIS RIVER NEAR KIRBYS CORNER	1826	SUP	2	Val	189	539
02CA007	THESSALON RIVER NEAR POPLAR DALE	255	HUR	2	Val	211	483

Gauge ID	Name	Derived Area [km^2]	Region	Obj.	Cal/ Val	Elev. [m]	Avg. Annual Runoff [mm]
02CB003	AUBINADONG RIVER ABOVE SESABIC CREEK	1481	HUR	1,2	Cal	347	362
02CC005	LITTLE WHITE RIVER NEAR BELLINGHAM	1954	HUR	2	Cal	218	406
02CD001	SERPENT RIVER AT HIGHWAY NO. 17	1252	HUR	2	Cal	178	464
02CE002	AUX SABLES RIVER AT MASSEY	1276	HUR	2	Val	170	420
02CF007	WHITSON RIVER AT CHELMSFORD	244	HUR	1,2	Cal	265	347
02CF010	ONAPING RIVER NEAR LEVACK	1651	HUR	2	Val	306	221
02CF011	VERMILION RIVER NEAR VAL CARON	701	HUR	1,2	Val	278	398
02DB005	WANAPITEI RIVER NEAR WANUP	3318	HUR	2	Cal	220	375
02DC012	STURGEON RIVER AT UPPER GOOSE FALLS	1281	HUR	1	Cal	256	406
02DD010	FRENCH RIVER AT DRY PINE BAY	18222	HUR	2	Cal	184	449
02EA005	NORTH MAGNETAWAN RIVER NEAR BURKS FALLS	325	HUR	1	Cal	309	593
02EA011	MAGNETAWAN RIVER NEAR BRITT	2633	HUR	2	Cal	178	509
02EA018	MAGNETAWAN RIVER NEAR EMSDALE	397	HUR	1	Val	332	586
02EB011	MOON RIVER AT HIGHWAY NO. 400	4527	HUR	2	Cal	209	182
02EB014	OXTONGUE RIVER NEAR DWIGHT	571	HUR	1	Cal	324	576
02EC002	BLACK RIVER NEAR WASHAGO	1451	HUR	1	Cal	208	514
02EC003	SEVERN RIVER AT SWIFT RAPIDS	5793	HUR	2	Cal	194	317
02EC011	BEAVER RIVER NEAR BEAVERTON	300	HUR	1	Val	227	347
02EC018	PEFFERLAW BROOK NEAR UDORA	348	HUR	1	Cal	229	290
02EC019	BLACK RIVER NEAR VANKOUGHNET	385	HUR	1	Val	275	588
02ED003	NOTTAWASAGA RIVER NEAR BAXTER	1295	HUR	1	Cal	189	270
02ED015	MAD RIVER AT AVENING	255	HUR	1	Cal	238	505
02ED024	North River at the Falls	249	HUR	1,2	Cal	171	459
02ED027	Nottawasaga River near Edenvale	2735	HUR	2	Val	163	324
02ED029	INNISFIL CREEK NEAR ALLISTON	457	HUR	1	Cal	202	226
02ED032	WILLOW CREEK NEAR MINESING	228	HUR	1	Val	192	351
02ED101	NOTTAWASAGA RIVER NEAR ALLISTON	337	HUR	1	Cal	222	315
02ED102	Boyne River at Earl Rowe Park	239	HUR	1	Cal	211	313
02FA001	SAUBLE RIVER AT SAUBLE FALLS	927	HUR	2	Cal	144	538

Gauge ID	Name	Derived Area [km^2]	Region	Obj.	Cal/ Val	Elev. [m]	Avg. Annual Runoff [mm]
02FA004	Sauble River at Allenford	349	HUR	1	Val	217	532
02FB009	BEAVER RIVER NEAR CLARKSBURG	586	HUR	2	Cal	225	510
02FB010	BIGHEAD RIVER NEAR MEAFORD	310	HUR	2	Cal	228	590
02FC001	SAUGEEN RIVER NEAR PORT ELGIN	3758	HUR	2	Cal	191	540
02FC016	SAUGEEN RIVER ABOVE DURHAM	348	HUR	1	Val	360	492
02FE008	Middle Maitland River near Belgrave	673	HUR	1	Val	309	489
02FE009	SOUTH MAITLAND RIVER AT SUMMERHILL	388	HUR	1	Cal	262	559
02FE010	BOYLE DRAIN NEAR ATWOOD	209	HUR	1	Val	353	456
02FE013	MIDDLE MAITLAND RIVER ABOVE ETHEL	402	HUR	1	Val	354	473
02FE015	Maitland River at Benmiller	2558	HUR	2	Cal	213	500
02FF002	AUSABLE RIVER NEAR SPRINGBANK	804	HUR	2	Val	201	407
02GA010	NITH RIVER NEAR CANNING	973	ERI	1	Cal	248	391
02GA018	NITH RIVER AT NEW HAMBURG	543	ERI	1	Cal	331	385
02GA038	NITH RIVER ABOVE NITHBURG	322	ERI	1	Cal	357	441
02GA047	SPEED RIVER AT CAMBRIDGE	782	ERI	1	Val	267	365
02GB001	GRAND RIVER AT BRANTFORD	5149	ERI	2	Cal	189	388
02GB007	FAIRCHILD CREEK NEAR BRANTFORD	389	ERI	1,2	Val	196	309
02GC002	KETTLE CREEK AT ST. THOMAS	334	ERI	1,2	Cal	205	355
02GC007	BIG CREEK NEAR WALSINGHAM	573	ERI	2	Cal	183	398
02GC010	BIG OTTER CREEK AT TILLSONBURG	393	ERI	1	Val	207	388
02GC026	Big Otter Creek near Calton	681	ERI	2	Val	181	418
02GD004	MIDDLE THAMES RIVER AT THAMESFORD	297	ERI	1	Cal	268	412
02GE003	THAMES RIVER AT THAMESVILLE	4498	ERI	2	Cal	171	416
02GE007	MCGREGOR CREEK NEAR CHATHAM	213	ERI	1,2	Val	161	333
02GG002	Sydenham River near Alvinston	712	ERI	1	Cal	195	349
02GG003	Sydenham River at Florence	1155	ERI	1,2	Cal	176	336
02GG006	BEAR CREEK NEAR PETROLIA	238	ERI	1	Cal	194	365
02GG009	BEAR CREEK BELOW BRIGDEN	520	ERI	1,2	Cal	187	334
02GG013	BLACK CREEK NEAR BRADSHAW	219	ERI	1,2	Val	183	386

Gauge ID	Name	Derived Area [km^2]	Region	Obj.	Cal/ Val	Elev. [m]	Avg. Annual Runoff [mm]
02HA006	TWENTY MILE CREEK AT BALLS FALLS	320	ONT	1,2	Cal	149	353
02HA007	WELLAND RIVER BELOW CAISTOR CORNERS	205	ERI	2	Cal	176	345
02HB011	BRONTE CREEK NEAR ZIMMERMAN	221	ONT	2	Val	161	338
02HB029	CREDIT RIVER AT STREETSVILLE	792	ONT	1,2	Val	149	369
02HC003	HUMBER RIVER AT WESTON	834	ONT	2	Cal	120	287
02HC024	DON RIVER AT TODMORDEN	344	ONT	2	Cal	76	428
02HC025	HUMBER RIVER AT ELDER MILLS	291	ONT	1	Val	158	275
02HC030	ETOBICOKE CREEK BELOW QUEEN ELIZABETH HIGHWAY	192	ONT	1,2	Cal	91	427
02HC049	Duffins Creek at Ajax	284	ONT	1,2	Cal	76	354
02HD012	Ganaraska River above Dale	243	ONT	1,2	Cal	105	435
02HF002	GULL RIVER AT NORLAND	1730	ONT	2	Val	269	517
02HF003	BURNT RIVER NEAR BURNT RIVER	1270	ONT	2	Cal	265	495
02HJ003	OUSE RIVER NEAR WESTWOOD	292	ONT	2	Val	189	339
02HK003	CROWE RIVER AT MARMORA	1892	ONT	2	Cal	158	402
02HL001	MOIRA RIVER NEAR FOXBORO	2673	ONT	2	Cal	84	386
02HL003	BLACK RIVER NEAR ACTINOLITE	413	ONT	1	Cal	155	415
02HL004	SKOOTAMATTA RIVER NEAR ACTINOLITE	672	ONT	1	Cal	150	397
02HL005	MOIRA RIVER NEAR DELORO	308	ONT	1	Cal	189	431
02HL008	CLARE RIVER NEAR BOGART	315	ONT	1	Val	142	414
02HM003	SALMON RIVER NEAR SHANNONVILLE	989	ONT	2	Cal	83	420
02HM007	Napanee River at Camden East	782	ONT	2	Cal	119	416
02HM010	SALMON RIVER AT TAMWORTH	588	ONT	1	Val	154	414
02JB013	KINOJEVIS (RIVIERE) A 0.3 KM EN AMONT DU PONT-ROUTE A CLERICY	2531	OTT	1	Val	262	478
02JC008	BLANCHE RIVER ABOVE ENGLEHART	1801	OTT	1	Cal	204	425
02JE027	AMABLE DU FOND RIVER AT KIOSK	706	OTT	1	Val	291	504
02KC018	INDIAN RIVER AT PEMBROKE	497	OTT	1	Val	120	376
02KD002	YORK RIVER NEAR BANCROFT	846	OTT	1	Val	335	513
02KF005	OTTAWA RIVER AT BRITANNIA	89925	OTT	2	Val	53	446

Gauge ID	Name	Derived Area [km^2]	Region	Obj.	Cal/ Val	Elev. [m]	Avg. Annual Runoff [mm]
02KF011	CARP RIVER NEAR KINBURN	266	OTT	1	Cal	87	389
02LA004	RIDEAU RIVER AT OTTAWA	4058	OTT	2	Cal	45	380
02LA006	KEMPTVILLE CREEK NEAR KEMPTVILLE	415	OTT	2	Cal	93	395
02LA007	JOCK RIVER NEAR RICHMOND	456	OTT	1,2	Cal	80	400
02LB005	SOUTH NATION RIVER NEAR PLANTAGENET SPRINGS	3819	OTT	2	Cal	33	430
02LB007	SOUTH NATION RIVER AT SPENCERVILLE	277	OTT	1	Val	89	435
02LB008	BEAR BROOK NEAR BOURGET	427	OTT	1	Cal	55	461
02LB032	RIGAUD RIVER NEAR ST. EUGENE	301	OTT	1,2	Val	42	621
02LC008	NORD (RIVIERE DU) A 4.8 KM EN AMONT DU PONT DU C.N. A SAINT-JEROME	1161	OTT	2	Val	88	684
02LC029	ROUGE (RIVIERE) EN AMONT DE LA CHUTE MCNEIL	5530	OTT	2	Val	131	631
02LD005	PETITE NATION (RIVIERE DE LA) AU PONT A 1.6 KM EN AMONT DE RIPON	1254	OTT	1,2	Val	158	559
02LE024	LIEVRE (RIVIERE DU) A 2.2 KM EN AMONT DU PONT-ROUTE 311 A LAC-SAINT-PAUL	4568	OTT	2	Val	240	594
02LE025	KIAMIKA (RIVIERE) A CHUTE-SAINT-PHILIPPE	882	OTT	2	Val	239	602
02LG005	GATINEAU (RIVIERE) AUX RAPIDES CEIZUR	6923	OTT	1,2	Val	221	572
04010500	PIGEON RIVER AT MIDDLE FALLS NR GRAND PORTAGE MN	1561	SUP	2	Cal	238	265
04015330	KNIFE RIVER NEAR TWO HARBORS MN	217	SUP	1,2	Val	183	350
04024000	ST. LOUIS RIVER AT SCANLON MN	8964	SUP	2	Cal	301	233
04024430	NEMADJI RIVER NEAR SOUTH SUPERIOR WI	1127	SUP	1,2	Cal	188	296
04027000	BAD RIVER NEAR ODANAH WI	1527	SUP	1,2	Cal	222	356
04027500	WHITE RIVER NEAR ASHLAND WI	735	SUP	2	Cal	226	302
04029990	Montreal River at Saxon Falls near Saxon WI	693	SUP	2	Cal	282	426
04031000	BLACK R NR BESSEMER MI	510	SUP	1,2	Cal	357	427
04040000	ONTONAGON R NR ROCKLAND MI	3491	SUP	2	Cal	200	315
04040500	STURGEON RIVER NEAR SIDNAW MI	477	SUP	1	Cal	378	392
04041500	STURGEON RIVER NEAR ALSTON MI	904	SUP	2	Val	233	380
04044724	Au Train River at Forest Lake MI	233	SUP	2	Cal	182	422

Gauge ID	Name	Derived Area [km^2]	Region	Obj.	Cal/ Val	Elev. [m]	Avg. Annual Runoff [mm]
04045500	TAHQUAMENON RIVER NR TAHQUAMENON PARADISE MI	2007	SUP	1,2	Cal	186	383
04056500	MANISTIQUE RIVER NR MANISTIQUE MI	2907	MIC	1,2	Cal	189	440
04057510	STURGEON RIVER NR NAHMA JUNCTION MI	516	MIC	1,2	Val	190	330
04059000	ESCANABA RIVER AT CORNELL MI	2265	MIC	2	Cal	221	272
04059500	FORD RIVER NR HYDE MI	1217	MIC	1,2	Cal	198	249
04066500	PIKE RIVER AT AMBERG WI	644	MIC	1	Cal	265	256
04067500	MENOMINEE RIVER NEAR MCALLISTER WI	10095	MIC	2	Cal	175	267
04067958	Peshtigo River near Wabeno WI	1217	MIC	1	Cal	306	277
04069500	PESHTIGO RIVER AT PESHTIGO WI	2810	MIC	2	Val	172	263
04071765	Oconto River near Oconto WI	2407	MIC	2	Cal	170	242
04072150	Duck Creek near Howard WI	271	MIC	2	Cal	169	199
04074950	WOLF RIVER AT LANGLADE WI	1206	MIC	1	Cal	383	290
04084500	FOX R AT RAPIDE CROCHE DAM NEAR WRIGHTSTOWN WI	15686	MIC	2	Val	152	258
04085200	KEWAUNEE RIVER NEAR KEWAUNEE WI	247	MIC	1,2	Cal	181	205
04085427	MANITOWOC RIVER AT MANITOWOC WI	1355	MIC	2	Cal	169	219
04086000	SHEBOYGAN RIVER AT SHEBOYGAN WI	1109	MIC	2	Cal	171	275
04087000	MILWAUKEE RIVER AT MILWAUKEE WI	1752	MIC	2	Cal	153	287
04087120	MENOMONEE RIVER AT WAUWATOSA WI	334	MIC	2	Val	186	350
04087240	ROOT RIVER AT RACINE WI	538	MIC	2	Cal	159	311
04093000	DEEP RIVER AT LAKE GEORGE OUTLET AT HOBART IN	346	MIC	2	Cal	155	421
04096015	Galien River near Sawyer MI	199	MIC	2	Val	165	380
04101500	ST. JOSEPH RIVER AT NILES MI	9962	MIC	2	Cal	188	355
04101800	DOWAGIAC RIVER AT SUMNERVILLE MI	635	MIC	2	Cal	177	413
04102500	PAW PAW RIVER AT RIVERSIDE MI	940	MIC	2	Cal	178	424
04102700	SOUTH BRANCH BLACK RIVER NEAR BANGOR MI	237	MIC	2	Cal	189	452
04102776	Middle Branch Black River near South Haven MI	214	MIC	1,2	Val	186	446
04108660	Kalamazoo River at New Richmond MI	5536	MIC	2	Val	147	403
04119000	GRAND RIVER AT GRAND RAPIDS MI	13324	MIC	2	Cal	162	311
04121970	Muskegon River near Croton MI	5992	MIC	2	Cal	206	314

Gauge ID	Name	Derived Area [km^2]	Region	Obj.	Cal/ Val	Elev. [m]	Avg. Annual Runoff [mm]
04122200	WHITE RIVER NEAR WHITEHALL MI	1090	MIC	1,2	Cal	187	387
04122500	PERE MARQUETTE RIVER AT SCOTTVILLE MI	1805	MIC	1,2	Cal	153	393
04124000	Manistee River near Sherman MI	2297	MIC	1	Cal	251	434
04125460	Pine River at High School Bridge nr Hoxeyville MI	645	MIC	1	Cal	255	415
04125550	Manistee River near Wellston MI	3586	MIC	2	Val	200	421
04126740	Platte River at Honor MI	341	MIC	2	Cal	167	346
04126970	Boardman R above Brown Bridge Road nr Mayfield MI	385	MIC	1,2	Cal	249	295
04133501	Thunder Bay River at Herron Road near Bolton MI	1461	HUR	2	Val	196	280
04136000	Au Sable River near Red Oak MI	2937	HUR	1	Cal	293	245
04137500	Au Sable River near Au Sable MI	4622	HUR	2	Cal	175	265
04142000	RIFLE RIVER NEAR STERLING MI	795	HUR	2	Cal	205	361
04159492	Black River near Jeddo MI	1147	ERI	2	Cal	195	225
04159900	Mill Creek near Avoca MI	470	ERI	2	Val	212	214
04160600	BELLE RIVER AT MEMPHIS MI	383	ERI	2	Cal	218	247
04165500	CLINTON RIVER AT MORAVIAN DRIVE AT MT. CLEMENS MI	1852	ERI	2	Cal	168	320
04166500	RIVER ROUGE AT DETROIT MI	479	ERI	2	Cal	186	294
04168400	Lower River Rouge at Dearborn MI	272	ERI	2	Val	173	396
04174500	HURON RIVER AT ANN ARBOR MI	1826	ERI	2	Cal	236	255
04176500	RIVER RAISIN NEAR MONROE MI	2822	ERI	2	Cal	177	280
04177000	OTTAWA RIVER At UNIVERSITY OF TOLEDO TOLEDO OH	336	ERI	2	Cal	178	297
04185000	Tiffin River at Stryker OH	988	ERI	1	Cal	203	324
04193500	MAUMEE R AT WATERVILLE OH	16393	ERI	2	Cal	162	345
04195500	PORTAGE R AT WOODVILLE OH	1099	ERI	2	Val	181	365
04196800	Tymochtee Creek at Crawford OH	663	ERI	1	Cal	215	369
04197100	Honey Creek at Melmore OH	390	ERI	1	Cal	242	379
04198000	SANDUSKY R NR FREMONT OH	3122	ERI	2	Cal	184	351
04199000	HURON R AT MILAN OH	915	ERI	2	Cal	180	383
04199500	VERMILION R NR VERMILION OH	663	ERI	2	Cal	186	404
04200500	BLACK R AT ELYRIA OH	983	ERI	2	Cal	174	380

Gauge ID	Name	Derived Area [km^2]	Region	Obj.	Cal/ Val	Elev. [m]	Avg. Annual Runoff [mm]
04201500	ROCKY R NR BEREA OH	696	ERI	2	Val	206	501
04208000	CUYAHOGA R AT INDEPENDENCE OH	1990	ERI	2	Cal	181	538
04209000	CHAGRIN R AT WILLOUGHBY OH	622	ERI	2	Val	195	639
04212100	GRAND R NR PAINESVILLE OH	1822	ERI	2	Cal	176	535
04213000	CONNEAUT C AT CONNEAUT OH	497	ERI	1,2	Cal	160	608
04213500	CATTARAUGUS CREEK AT GOWANDA NY	1169	ERI	2	Cal	233	640
04214500	BUFFALO CREEK AT GARDENVILLE NY	382	ERI	2	Val	191	548
04215000	CAYUGA CREEK NR LANCASTER NY	255	ERI	2	Cal	212	570
04215500	CAZENOVIA CREEK AT EBENEZER NY	347	ERI	2	Cal	189	702
04218000	TONAWANDA CREEK AT RAPIDS NY	895	ERI	2	Cal	156	476
04218518	ELLICOTT CREEK BELOW WILLIAMSVILLE NY	232	ERI	2	Cal	160	636
04220045	OAK ORCHARD CREEK NEAR SHELBY NY	385	ONT	2	Val	175	366
04221000	GENESEE RIVER AT WELLSVILLE NY	762	ONT	1	Cal	443	479
04224775	CANASERAGA CREEK ABOVE DANSVILLE NY	248	ONT	1	Cal	199	396
04231600	GENESEE RIVER AT FORD STREET BRIDGE ROCHESTER NY	6338	ONT	2	Cal	125	417
0423205010	Irondequoit Cr abv Blossom Rd nr Rochester NY	363	ONT	2	Cal	83	354
04249000	OSWEGO RIVER AT LOCK 7 AT OSWEGO NY	13221	ONT	2	Val	51	522
04250200	Salmon River at Pineville NY	644	ONT	2	Cal	139	1189
04250750	SANDY CREEK NEAR ADAMS NY	373	ONT	2	Val	159	740
04256000	INDEPENDENCE RIVER AT DONNATTSBURG NY	229	ONT	1,2	Cal	306	838
04258000	Beaver River at Croghan NY	754	ONT	2	Cal	248	814

S.5 Common set of donor basins

375   The seven locally calibrated models (LBRM-CC-lumped, HYMOD2-lumped, GR4J-lumped, HMETS-lumped, Blended-lumped, Blended-Raven, and VIC-Raven) needed to specify a so-called donor basin for each of the 71 validation locations. The donor basin is a basin that was calibrated. The calibrated parameter set of the donor basin is then used to run the model deriving streamflow at the validation location. All models used the mapping of donor basins given in Table S16. The mapping approach is explained in the main manuscript (Sect. 2.4).

Table S16: **Donor basin mapping.** The Table specifies the donor basin for each validation location. These donor basins are used by all locally calibrated models. The table also specifies the reason for the donor basin to be selected.

Validation Basin	Donor Basin	Reason Selected
02AB006	02AB017	closest basin centroids
02AB021	02AC001	closest basin centroids
02AD012	02AD010	nested basins
02AE001	02AD010	closest basin centroids
02BA006	02BA003	closest basin centroids
02BC004	02BB004	closest basin centroids
02BC006	02BB004	closest basin centroids
02BD007	02BD002	closest basin centroids
02BE002	02BF001	closest basin centroids
02BF014	02BF002	nested basins
02CA007	02BF002	closest basin centroids
02CE002	02CD001	closest basin centroids
02CF010	02DB005	closest basin centroids
02CF011	02CF007	closest basin centroids
02EA018	02EA011	nested basins
02EC011	02EC003	nested basins
02EC019	02EC002	nested basins
02ED027	02ED003	nested basins
02ED032	02ED024	closest basin centroids
02FA004	02FA001	nested basins
02FC016	02FC001	nested basins
02FE008	02FE015	nested basins
02FE010	02FE015	nested basins
02FE013	02FE015	nested basins
02FF002	02GG002	closest basin centroids
02GA047	02GB001	nested basins

Validation Basin	Donor Basin	Reason Selected
02GB007	02GC007	closest basin centroids
02GC010	02GC007	closest basin centroids
02GC026	02GC007	closest basin centroids
02GE007	02GG003	closest basin centroids
02GG013	02GG009	closest basin centroids
02HB011	02HC030	closest basin centroids
02HB029	02HC030	closest basin centroids
02HC025	02HC003	nested basins
02HF002	02EB014	closest basin centroids
02HJ003	02HK003	closest basin centroids
02HL008	02HL001	nested basins
02HM010	02HL004	closest basin centroids
02JB013	02JC008	closest basin centroids
02JE027	02EA005	closest basin centroids
02KC018	02HL004	closest basin centroids
02KD002	02HF003	closest basin centroids
02KF005	02JC008	nested basins
02LB007	02LB005	nested basins
02LB032	02LB005	closest basin centroids
02LC008	02LB008	closest basin centroids
02LC029	02LB008	closest basin centroids
02LD005	02LB008	closest basin centroids
02LE024	02LB008	closest basin centroids
02LE025	02LB008	closest basin centroids
02LG005	02LB008	closest basin centroids
04015330	04024000	closest basin centroids
04041500	04040500	nested basins
04057510	04044724	closest basin centroids
04069500	04067958	nested basins
04084500	04074950	nested basins
04087120	04087240	closest basin centroids
04096015	04101800	closest basin centroids
04102776	04102700	closest basin centroids
04108660	04102700	closest basin centroids
04125550	04124000	nested basins

Validation Basin	Donor Basin	Reason Selected
04133501	04137500	closest basin centroids
04159900	04160600	closest basin centroids
04168400	04176500	closest basin centroids
04195500	04198000	closest basin centroids
04201500	04200500	closest basin centroids
04209000	04208000	closest basin centroids
04214500	04215000	closest basin centroids
04220045	04218000	closest basin centroids
04249000	0423205010	closest basin centroids
04250750	04250200	closest basin centroids

380

References

Allen, R. G., Pereira, L. S., Raes, D., Smith, M., et al.: Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, Fao, Rome, 300, D05 109, 1998.

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, Journal of Hydrometeorology, 12, 508–530, 2011.

Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, Water Resources Research, 36, 3663–3674, 2000.

Chlumsky, R., Mai, J., Craig, J. R., and Tolson, B. A.: Simultaneous Calibration of Hydrologic Model Structure and Parameters Using a Blended Model, Water Resources Research, 57, e2020WR029 229, 2021.

Craig, J. R.: Raven: User's and Developer's Manual v3.5, accessed: 2022-02-02, 2022.

Craig, J. R., Brown, G., Chlumsky, R., Jenkinson, R. W., Jost, G., Lee, K., Mai, J., Serrer, M., Sgro, N., Shafii, M., Snowdon, A. P., and Tolson, B. A.: Flexible watershed simulation with the Raven hydrological modelling framework, Environmental Modelling & Software, 129, 104 728, 2020.

Crowley II, T. E.: Great Lakes basins (USA-Canada) runoff modeling, Journal of Hydrology, 64, 135–158, 1983.

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, Journal of Hydrology, 521, 374–388, 2015.

Durnford, D., Carrera, M., Dupont, F., Deacu, D., Gaborit, É., Garnaud, C., Fortin, V., Bélair, S., Lespinas, F., Bilodeau, B., Khedhaouiria, D., Gauthier, N., Roy, G., Matte, P., Vionnet, V., Badawy, B., Liu, X., Bekcic, B., Shin, L., Champoux, O., Abahamowicz, M., Keita, S., Morales-Marin, L., Morin, J., Hata, Y., and Martinez, Y.: Hydrological Prediction Systems at Environment and Climate Change Canada, in: American Meteorological Society 101st Annual Meeting, pp. 586–597, 2021.

Fry, L. M., Apps, D., and Gronewold, A. D.: Operational Seasonal Water Supply and Water Level Forecasting for the Laurentian Great Lakes, Journal of Water Resources Planning and Management, 146, 04020 072, 2020.

Gaborit, É., Muschalla, D., Vallet, B., Vanrolleghem, P. A., and Anctil, F.: Improving the performance of stormwater detention basins by real-time control using rainfall forecasts, Urban Water Journal, 10, 230–246, 2013.

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, Canadian Journal of Earth Sciences, 52, 990–999, 2015.

Gaborit, É., Fortin, V., Xu, X., Seglenieks, F., Tolson, B., Fry, L. M., Hunter, T., Anctil, F., and Gronewold, A. D.: A hydrological prediction system based on the SVS land-surface scheme: efficient calibration of GEM-Hydro for streamflow simulation over the Lake Ontario basin, Hydrology and Earth System Sciences, 21, 4825–4839, 2017.

Gao, H., Tang, Q., Shi, X., Zhu, C., Bohn, T., Su, F., Pan, M., Sheffield, J., Lettenmaier, D., and Wood, E.: Water Budget Record from Variable Infiltration Capacity (VIC) Model, in: Algorithm Theoretical Basis Document for Terrestrial Water Cycle Data Records, pp. 120–173, 2010.

Gers, F. A., Schmidhuber, J., and Cummins, F.: Learning to forget: continual prediction with LSTM, IET Conference Proceedings, pp. 850–855, 1999.

Gronewold, A. D., Hunter, T., Allison, J., Fry, L. M., Kompoltowicz, K. A., Bolinger, R. A., and Pei, L.: Project Documentation Report for Great Lakes seasonal and inter-annual water supply forecasting improvements project Phase I: Research and Development., Tech. rep., NOAA-GLERL, Ann Arbor, MI, https://www.glerl.noaa.gov/pubs/fulltext/2018/20180020.pdf, 2017.

Han, M.: BasinMaker GIS Toolbox v1.0, https://github.com/dustming/basinmaker/releases/tag/v1.0, accessed: 2022-01-19, 2021.

Han, M., Mai, J., Tolson, B. A., Craig, J. R., Gaborit, É., Liu, H., and Lee, K.: Subwatershed-based lake and river routing products for hydrologic and land surface models applied over Canada, Canadian Water Resources Journal, 0, 1–15, 2020.

Han, M., Mai, J., Tolson, B. A., Craig, J. R., and Shen, H.: BasinMaker GIS Toolbox website, http://hydrology.uwaterloo.ca/basinmaker/, accessed: 2022-01-19, 2021.

Hargreaves, G. H. and Samani, Z. A.: Reference crop evapotranspiration from temperature, Applied Engineering in Agriculture, 1, 96–99, 1985.

Hastie, T., Tibshirani, R., and Friedman, J.: Model Assessment and Selection, pp. 219–259, Springer New York, New York, NY, https://doi.org/10.1007/978-0-387-84858-7_7, https://doi.org/10.1007/978-0-387-84858-7_7, 2009.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Improving neural networks by preventing co-adaptation of feature detectors, CoRR, abs/1207.0580, http://arxiv.org/abs/1207.0580, 2012.

King, F., Erler, A. R., Frey, S. K., and Fletcher, C. G.: Application of machine learning techniques for regional bias correction of snow water equivalent estimates in Ontario, Canada, Hydrol. Earth Syst. Sci., 24, 4887–4902, https://doi.org/https://doi.org/10.5194/hess-24-4887-2020, 2020.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, in: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, edited by Bengio, Y. and LeCun, Y., http://arxiv.org/abs/1412.6980, 2015.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrology and Earth System Sciences, 23, 5089–5110, https://doi.org/10.5194/hess-23-5089-2019, 2019.

Kratzert, F., Gauch, M., Nearing, G. S., and Klotz, D.: NeuralHydrology — A Python library for Deep Learning research in hydrology, Journal of Open Source Software, 7, 4050, 2022.

Li, B., Rodell, M., and Famiglietti, J. S.: Groundwater variability across temporal and spatial scales in the central and northeastern U.S., Journal of Hydrology, 525, 769–780, 2015.

Lofgren, B. M. and Rouhana, J.: Physically Plausible Methods for Projecting Changes in Great Lakes Water Levels under Climate Change Scenarios, Journal of Hydrometeorology, 17, 2209–2223, 2016.

Lofgren, B. M., Hunter, T. S., and Wilbarger, J.: Effects of using air temperature as a proxy for potential evapotranspiration in climate change scenarios of Great Lakes basin hydrology, Journal of Great Lakes Research, 37, 744–752, 2011.

Mai, J., Craig, J. R., and Tolson, B. A.: Simultaneously determining global sensitivities of model parameters and model structure, Hydrology and Earth System Sciences, 24, 5835–5858, 2020.

Mai, J., Tolson, B. A., Shen, H., Gaborit, É., Fortin, V., Gasset, N., Awoye, H., Stadnyk, T. A., Fry, L. M., Bradley, E. A., Seglenieks, F., Temgoua, A. G. T., Princz, D. G., Gharari, S., Haghnegahdar, A., Elshamy, M. E., Razavi, S., Gauch, M., Lin, J., Ni, X., Yuan, Y., McLeod, M., Basu, N. B., Kumar, R., Rakovec, O., Samaniego, L., Attinger, S., Shrestha, N. K., Daggupati, P., Roy, T., Wi, S., Hunter, T., Craig, J. R., and Pietroniro, A.: Great Lakes Runoff Intercomparison Project Phase 3: Lake Erie (GRIP-E), Journal of Hydrologic Engineering, 26, 05021 020, 2021.

Mai, J., Craig, J. R., Tolson, B. A., and Arsenault, R.: The sensitivity of simulated streamflow to individual hydrologic processes across North America, Nature Communications, 13, 455, 2022.

455 Martel, J.-L., Demeester, K., Brissette, F., Poulin, A., and Arsenault, R.: HMETS—A Simple and Efficient Hydrology Model for Teaching Hydrological Modelling, Flow Forecasting and Climate Change Impacts, International Journal of Engineering Education, 33, 1307–1316, 2017.

Martinec, J.: Snowmelt-runoff model for stream flow forecasts , Nordic Hydrology, 6, 145–154, 1975.

Mekis, E., Stewart, R. E., Theriault, J. M., Kochtubajda, B., Bonsal, B. R., and Liu, Z.: Near-0°C surface temperature and precipitation type
460 patterns across Canada, Hydrol. Earth Syst. Sci., 24, 1741–1761, https://doi.org/https://doi.org/10.5194/hess-24-1741-2020, 2020.

Moore, R. J.: The probability-distributed principle and runoff production at point and basin scales, Hydrological Sciences Journal, 30, 273–297, 1985.

Muñoz Sabater, J.: ERA5-Land hourly data from 1981 to present., https://doi.org/https://doi.org/10.24381/cds.e2161bac, Copernicus Climate Change Service (C3S) Climate Data Store (CDS) (Accessed on 04-10-2021), 2019.

465 Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrology and Earth System Sciences, 19, 209–223, https://doi.org/10.5194/hess-19-209-2015, https://hess.copernicus.org/articles/19/209/2015/, 2015.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279,
470 275–289, 2003.

Roy, T., Gupta, H. V., Serrat-Capdevila, A., and Valdes, J. B.: Using satellite-based evapotranspiration estimates to improve the structure of a simple conceptual rainfall–runoff model, Hydrology and Earth System Sciences, 21, 879–896, 2017.

Valayamkunnath, P., Barlage, M., Chen, F., Gochis, D. J., and Franz, K. J.: Mapping of 30-meter resolution tile-drained croplands using a geospatial modeling approach, Scientific Data, 7, 1–10, 2020.

475 Valéry, A., Andréassian, V., and Perrin, C.: "As simple as possible but not simpler": What is useful in a temperature-based snow-accounting routine? Part 1 - Comparison of six snow accounting routines on 380 catchments, Journal of Hydrology, 517, 1166–1175, 2014.

Vionnet, V., Mortimer, C., Brady, M., Arnal, L., and Brown, R.: Canadian historical Snow Water Equivalent dataset (CanSWE, 1928–2020), Earth System Science Data, 13, 4603–4619, 2021.