





The sensitivity of simulated streamflow to individual hydrologic processes across North America

Juliane Mai ¹✉, James R. Craig ¹, Bryan A. Tolson ¹ & Richard Arsenault ²

Streamflow sensitivity to different hydrologic processes varies in both space and time. This sensitivity is traditionally evaluated for the parameters specific to a given hydrologic model simulating streamflow. In this study, we apply a novel analysis over more than 3000 basins across North America considering a blended hydrologic model structure, which includes not only parametric, but also structural uncertainties. This enables seamless quantification of model process sensitivities and parameter sensitivities across a continuous set of models. It also leads to high-level conclusions about the importance of water cycle components on streamflow predictions, such as quickflow being the most sensitive process for streamflow simulations across the North American continent. The results of the 3000 basins are used to derive an approximation of sensitivities based on physiographic and climatologic data without the need to perform expensive sensitivity analyses. Detailed spatio-temporal inputs and results are shared through an interactive website.

¹Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, ON, Canada. ²Department of Construction Engineering, École de technologie supérieure, Montreal, QC, Canada. ✉email: juliane.mai@uwaterloo.ca

Hydrologic models are widely used in applications that are important for society such as flood prediction^{1–6}, drought monitoring^{7–10}, infrastructure design^{11–13}, and reservoir management^{14–16}. This wide variety of such applications, coupled with the diversity of climatic and physiographic regions and the underlying complexity of hydrologic processes is leading to increasing complexity among these models^{17–19}. Further developments and improvements of hydrologic models are essential to advance the understanding of hydrologic processes and ensure greater model realism^{20–23}. One way such improvements can be ensured is by carrying out model evaluations taking advantage of information theory and newly available datasets^{24–29}. Sensitivity analysis (SA) is a well-established tool to guide such model assessments³⁰, navigate model development³¹, and identify the most critical relationships within a system^{32–34}. SA is based on large sets of model runs identifying the most sensitive parameters of a model and is thus a general method that can be applied to any kind of model that contains unknown parameter estimates^{35–37}. Note that parameters can be (traditional) model parameters, multiplicative factors to perturb input forcings, or parameters to weight between different options, among others.

Notwithstanding the repute of SA as a tool, there are several challenges limiting the transferability and insights of individual analyses. Four such challenges, here highlighted for hydrologic applications, are:

Model parameters only: SAs traditionally only estimate the sensitivities of model parameters on streamflow^{35,38} or sensitivity indices of parameters on components or processes of the water cycle³⁴, rather than quantifying the sensitivity of streamflow to hydrologic processes, which limits insights in process understanding. Parameter-based analyses are model-specific and rarely lead to conclusions that can be transferred to other models.

Dependence on location: SAs are based on thousands of model runs, which makes them computationally very expensive. They are therefore usually only carried out for individual locations^{39–42}, and this limits the transferability of the obtained results to other locations.

Dependence on model structure: SAs are generally performed for individual models, which further limits the generality of conclusions^{30,33,43}.

Data sharing and re-usability: In-depth SA results, especially when applied to multiple locations using complex models, are usually not shared in an easily accessible way due to the amount and complexity of the data, which makes it challenging to obtain or compile information of interest for further usage^{44–46}.

This work applies the extended Sobol' Sensitivity Analysis (xSSA) method of Mai et al.⁴⁷ to a set of more than 3000 modelled locations across North America. The novelty of the xSSA method is that it generates process sensitivities in addition to the traditionally derived parameter sensitivities (addressing limitation 1). The continental-scale deployment allows for conclusions that hold over large domains (addressing limitation 2). The method is applied to the "Blended Model" for streamflow simulation introduced by Mai et al.⁴⁷, which enables a seamless analysis of model parameters and model structures, reducing the dependence of the results on specific modeling choices (addressing limitation 3). The input data, model setups, and SA results are shared on a map-based interactive website for members of the hydrologic research community to browse, explore, download and use for their specific regions of interest (addressing limitation 4).

Results

Figure 1 provides an overview of the analyses performed in this study. The HYSETS database⁴⁸ is screened for watersheds with adequate overall data availability and catchment size, and the

blended hydrological model is developed, calibrated, and tested in validation. Models with adequate performance in the calibration period are subjected to xSSA analyses, enabling the deduction of functional relationships between basin attributes and the sensitivity of hydrologic processes at any location.

Preliminary calibration and validation of the blended model.

Figure 2 shows the results of the preliminary basin-wise calibration of the 3826 basins with enough observed data in the calibration period (January 1991 to December 2010), as well as the performance of the 3005 calibrated basins which have enough data available during the validation period (January 1971 to December 1990). The median daily streamflow Nash-Sutcliffe efficiency (NSE) is 0.73 in calibration and 0.64 in validation. This is comparable to the performance of other models applied across the continental US (CONUS). For instance, Rakovec et al.⁴⁹ reported a median NSE of 0.72/0.66 (calibration/validation) for the mHM model^{50,51} over 492 CONUS basins. Mizukami et al.⁵² and Rakovec et al.⁴⁹ reported median NSEs of 0.61/0.57 (calibration/validation) for the VIC model^{53,54} applied to the same basins, and⁵⁵ reported median NSEs of 0.7 to 0.75/0.6 to 0.65 (calibration/validation) using the SAC-SMA/Snow-17 model^{56,57} over 671 CONUS basins, with variations due to meteorological forcings. The weaker performance of the blended model during validation over the high plains and desert southwest (seen in Fig. 2) is consistent with all the above-mentioned models. Regions with NSE performances lower than 0.5 are considered to be unreliable, and the basins are not included in the analyses to follow. The 3316 basins with an NSE of at least 0.5 are used going forward from here on. The selection of basins based on this threshold is applied for calibration performance rather than validation performance as the calibration period is the period used for the sensitivity analysis.

It is important to note that the results of this calibration exercise are used to (1) exclude clearly low quality models from further analysis and (2) to demonstrate the basic adequacy of the models for simulating streamflow over the range of simulated conditions. While a more elaborate calibration study may improve individual optimal model performances, it is unlikely to yield improved global sensitivity estimates. The exclusion of low quality models is admittedly not standard practice in sensitivity analyses. However, ensuring that models are able to represent physical processes is critical to confidently conclude on the spatial behavior of process sensitivities. Detailed results for all basins calibrated and validated, including the calibrated model setups, can be found on the website⁵⁸ associated with this publication.

Spatial variation in hydrologic process sensitivity. Figure 3 shows the variance-weighted total Sobol' sensitivity index ST_i^w , a metric representing the sensitivity of streamflow to variations of hydrologic processes. The variance weighting of timesteps for temporal aggregation is chosen to increase the importance of timesteps with high flows, which, in general, is a favored scheme in hydrologic applications^{33,59}. The total Sobol' index includes interactions between parameters and between processes. The results show clear spatial patterns of the importance of the hydrologic processes with regard to streamflow. The patterns are in agreement with hydrologic reasoning; for example, snow balance sensitivity is high in mountainous and northern regions and potential melt is only sensitive where snow occurs. The sensitivity analysis results were determined to be robust to changes in specified parameter ranges, with negligible (<0.0135) changes to process sensitivities based an analysis using a subset of 150 randomly selected basins (results not shown).

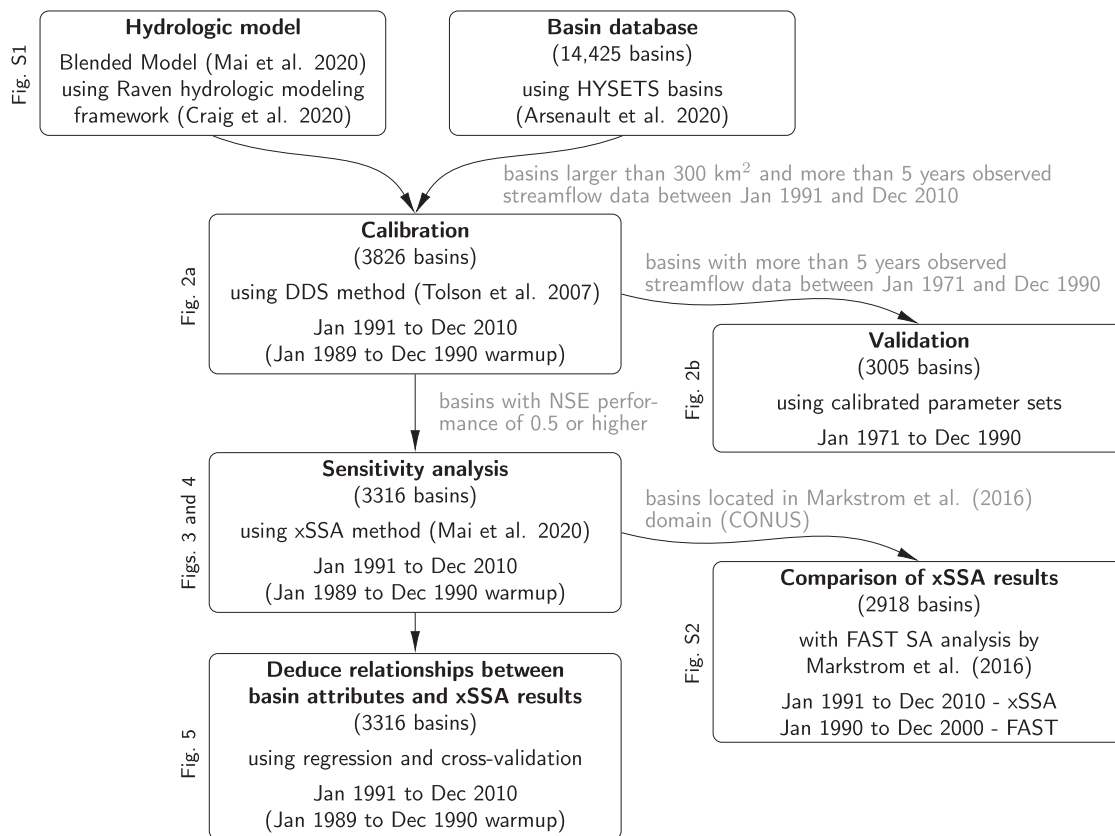


Fig. 1 Flowchart of experiments and analyses. All experiments performed in this work are listed, including the number of basins on which the analyses are based, the methods used, and the time periods over which the analyses are performed. The reasons for reducing the number of basins from the original 14,425 in the HYSETS database to 3316 for the xSSA sensitivity analysis and the deduction of relationships between basin characteristics and xSSA sensitivity results are added as gray labels to the arrows. The figures displaying the main results of each analysis are added as labels to the left of each box describing the analysis. The methods and results of the comparison of the xSSA results with another study can be found in the Supplementary Material (Fig. S2).

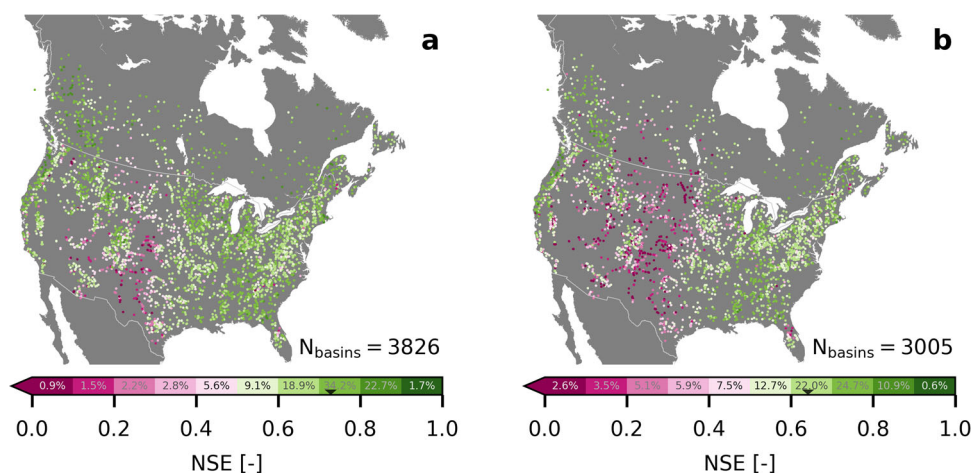


Fig. 2 Preliminary calibration and validation results of blended model. The performance with respect to the daily Nash-Sutcliffe efficiency (NSE) of the blended model during (a) calibration period (January 1991 to December 2010) and (b) validation period (January 1971 to December 1990). In total, 3826 basins with more than five years of streamflow observation data available during the calibration periods are calibrated. The 3005 basins that also had more than five years of data available during the validation period are validated. Each basin is represented on the map by its location of the streamflow gauge station (colored dots), while the color indicates the NSE performance. The distribution of basin performances is indicated on the colorbar. The black triangles (▼) in the colorbars mark the median NSE performance of 0.73 for calibration and 0.64 for validation. The 3316 basins with an NSE performance of 0.5 or higher during calibration are used for the remaining analyses of this study.

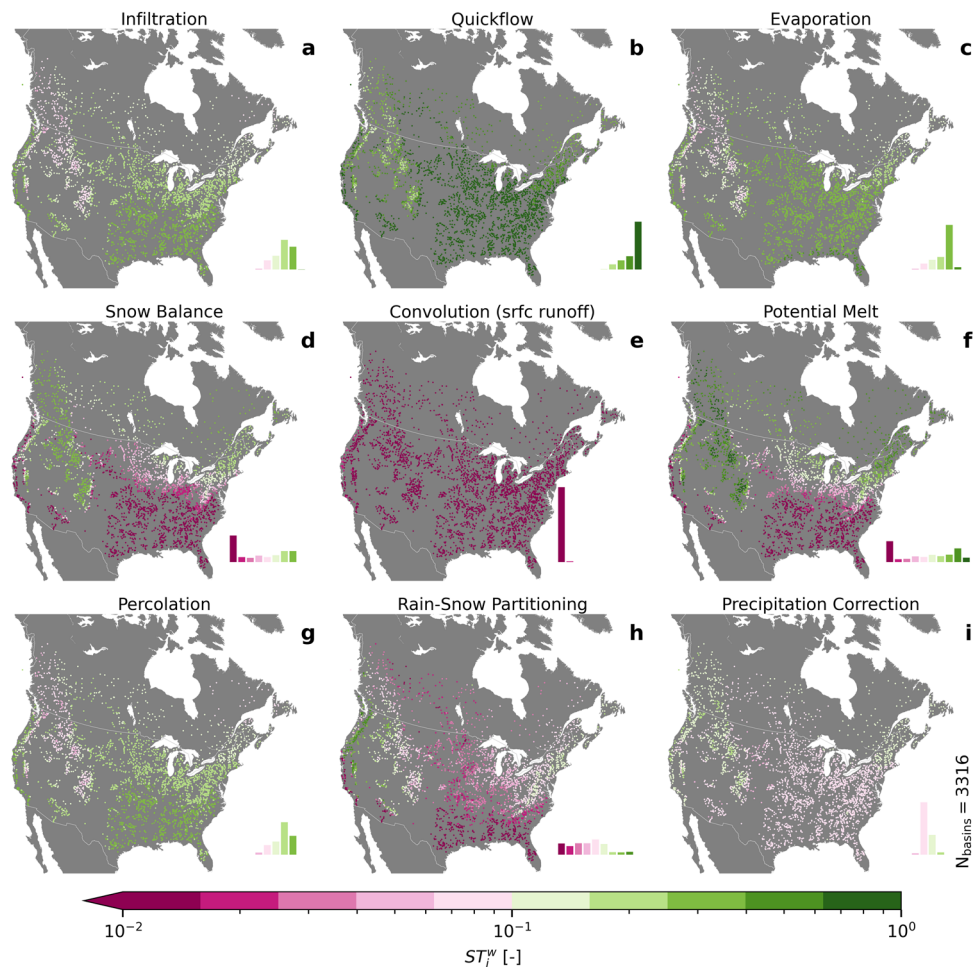


Fig. 3 Importance of hydrologic processes for streamflow simulations. The variance-weighted total Sobol' sensitivity ST_i^w regarding simulated streamflow for the nine hydrologic model components, i.e.: (a) infiltration, (b) quickflow, (c) evaporation, (d) snow balance, (e) convolution of surface runoff, (f) potential melt, (g) percolation, (h) rain-snow partitioning, and (i) precipitation correction. The results are shown for the 3316 basins that have a calibrated Nash-Sutcliffe efficiency equal to or better than 0.5 in the calibration period (January 1991 to December 2010). The two additional processes of Baseflow and Convolution (delayed runoff) are analyzed but not displayed as all basins show a sensitivity of less than 0.01 (lower limit of colorbar). Please note that the colorbar is logarithmic in order to allow for a better distinction of small sensitivity estimates. Green colors indicate a large importance of the respective process on streamflow simulations while pink colors indicate a weak impact. The location of the dots in each panel marks the location of the outlet of each basin, which coincides with a streamflow gauging station. The histograms of sensitivity values are added as an inset to each map. The histograms bins are the same as used for the colorbar.

Overall, quickflow (Fig. 3b) is the most important process for streamflow simulations, with a median ST_i^w of 0.736 across the 3316 basins analyzed. It exhibits large sensitivities, especially in the Eastern United States, which is covered mostly in temperate broadleaf and mixed forests. Infiltration, evaporation, and percolation (Fig. 3a, c, and g, respectively) are processes of secondary importance across North America, but especially in the Eastern US, with median ST_i^w of 0.300, 0.21, and 0.202, respectively. In the Rocky Mountains and in higher latitudes (starting over the Great Lakes), the model's potential melt, precipitation correction and snow balance components (Fig. 3f, i, and d) are the most important, with an overall median ST_i^w of 0.102, 0.087, and 0.047, based on the 3316 basins analyzed. On the West Coast, rain-snow partitioning (Fig. 3h) becomes more influential, with an overall median ST_i^w of 0.054. The convolution of surface runoff (Fig. 3e), a process which controls the timing (rather than magnitude) of flow, is found to be insensitive throughout, with a median ST_i^w of 0.003. The convolution of delayed runoff and baseflow are even less sensitive, with an ST_i^w below 0.01 in each basin. Both latter processes are thus not shown

in Fig. 3. It is not surprising that these three processes (convolution of surface and delayed runoff and baseflow) are less sensitive since the sensitivity analysis assesses the variability in streamflow magnitudes rather than its timing.

The lack of sensitivity studies conducted over large domains and the novelty of the sensitivity method presented by Mai et al.⁴⁷ in estimating sensitivities of processes rather than model parameters lead to challenges in comparing results to those of previous studies. However, a large-scale sensitivity study across the continental US was performed by Markstrom et al.³⁴ using the US Geological Survey's Precipitation-Runoff Modeling System (PRMS)⁶⁰. Their study derives the sensitivity of model parameters on eight hydrologic processes outputs. We compared the first-order mean (rather than time-dependent) sensitivities of runoff equivalent to S_i in the Markstrom et al.⁶⁰ study, and the results are presented in the Supplementary Material (Fig. S2). While the sensitivity metrics are (strictly speaking) not equivalent, a correlation test yields a Pearson correlation coefficient of 0.88, and the sensitivities exhibit similar spatial trends as those presented here. The analysis herein furthermore

represents an improvement upon the Markstrom et al.³⁴ study, as it provides insights averaged over several model structures, is based on time-dependent sensitivities, and includes sensitivity estimates accounting for parameter interactions (ST_i) besides the first-order effects (S_i).

The sensitivity of hydrologic processes at the continental scale has been highlighted here due to its novelty. However, the xSSA analysis does not only derive the sensitivity of the hydrologic processes but also derives the sensitivity of the parameters and process options of the blended model regarding the time series of simulated streamflow. The detailed results, including maps of average main and total Sobol' sensitivity indices (S_i^m and ST_i^m) and variance-weighted main and total Sobol' sensitivity indices (S_i^w and ST_i^w) for all parameters, process options, and processes, as well as the according summary plots, are available on the website associated with this publication⁶¹.

Regional variation in transient hydrologic process sensitivity.

Figure 4 shows the time-dependent sensitivities of each process clustered by similarity. The time-dependent total sensitivities ST_i regarding the simulated streamflow $Q(t)$ are averaged for each day of the year for the 20-year simulation period from January 1991 to December 2010. Each plot shown is a representative example basin of each clustered region (Fig. 4a, h). The regions are obtained by a c-means fuzzy clustering⁶² based on the three climate indicators, namely, aridity, seasonality, and fraction of precipitation as snow. The example basin is then selected as being the closest to the cluster centroid based on the climate indicators (not the spatial centroid). Although the clustering is unsupervised, the identified regions align well with known physiographic and ecological regions, and named here according to this alignment.

In all regions, quickflow (medium blue) is the most important process, as already deduced from the time-averaged sensitivities in Fig. 3. Evaporation (dark blue) is also important, but mostly during low-flow periods (consistent with low weights represented by the black line). Potential melt (i.e., incoming energy, in orange) and snow balance (i.e., snow ablation processes, in green) are mostly important during cold months, and especially during the freshet (if it exists). Baseflow (dark green) is only visible in some regions (Fig. 4c, e, f), and only during severe low-flow periods. The two convolution processes controlling the timing of runoff (light green and yellow) are important almost nowhere, except in regions with large freshet events in April and May (Fig. 4g, h). The overall results are consistent with the time-aggregated sensitivities. However, the variation in sensitivity of processes throughout the year yields more detailed insights into the seasonal variability in hydrological process influence across basins.

To elaborate, the prototypical responses in the Coastal and Interior Plains (Fig. 4a) and Arid Regions and Florida regions (Fig. 4b) are almost similar. The two regions show near-constant sensitivities throughout the year for almost all active processes (i.e., infiltration, quickflow, evaporation, potential melt, percolation, and rain-snow partitioning). Potential melt (orange) and rain-snow partitioning (medium red) exhibit an increased importance during the winter months (December to February and December to March), while no overall strong streamflow variability is present throughout the year (as seen from the timestep weight in black).

The prototypical basin for the Mediterranean California and Temperate Sierra region (Fig. 4c) shows a higher streamflow during winter months and an elevated importance of rain-snow partitioning (medium red) during that time period, while baseflow (dark green) becomes relevant during the low-flow periods.

The fourth cluster of mainly Temperate Broadleaf and Mixed Forests (Fig. 4d) region with mild freshet and regular mid-winter melt events is the first basin showing an impact of snow balance (medium green) on the streamflow simulations during the winter months (January to March).

The basin representing the Boreal Forest region (Fig. 4e) shows that baseflow (dark green) becomes important during the severe low-flow winter period (December to mid-February) and potential melt (orange), as well as snow balance (green), are highly elevated during the freshet (March to April).

The example basin for the Temperate Coniferous Forests region (Fig. 4f) has large amounts of precipitation during the winter and spring months (large weights between December and April), but almost none is snow (blue channel of RGB of climate index is small). The summer months (June to October) of this catchment are fairly dry (low weight equals low flow). This leads to two sensitivity regimes: during winter, snow balance (green), potential melt (orange) and rain-snow partitioning (medium red) are elevated, while during summer, evaporation (dark blue), percolation (light red) and baseflow (dark green) become more important. The importance of infiltration (light blue) and quickflow (medium blue), however, is almost constant throughout the entire year. The latter shows a slightly decreased sensitivity during the melt period (March and April) when other processes become more important.

The freshet, with its extreme high-flow periods (April to May), is even more pronounced for the two examples in the Strongly Seasonal and Snow-Dominated Regions (Fig. 4g) and Montane Cordillera (Fig. 4h) snow-dominated clusters. The latter has a longer flattened high-flow period, while the first generally peaks in April and early May. In both basins, snow balance (green) and potential melt (orange) have an increased sensitivity during the high-flow freshet period, while during the low-flow summer months, evaporation (dark blue) is gaining importance. In all cases, the results of the sensitivity analysis performed here are generally consistent with our hydrological expectations of these landform types.

The temporal sensitivity patterns are catchment-specific, and vary between basins within a region. The temporal sensitivities of the 3316 analyzed basins can be viewed on the interactive map on the website associated with this publication⁶³.

Estimating process sensitivity directly from basin attributes.

The sensitivities generated through xSSA were regressed against basin characteristics such as basin area and climatology (e.g., annual total precipitation based on the entire period of available forcings from 1950 to 2010), in order to assess whether sensitivities were readily determinable without the extensive analysis performed here. Figure 5 shows the predictability of the process sensitivities ST_i^w using these regressions, with each functional relationship for each process using one predictor unless the use of two predictors increased the adjusted coefficient of determination R_{adj}^2 by at least 0.05. The regression model selection was guided by performance in predictive mode assessed in cross-validation experiments. The regressed relationships are given in Table 1. The relationships are based on regressions using the entire set of 3316 basins (the skill of that regression is given as R_{adj}^2 in the table). To obtain a measure of how sensitive these functions are to the choice of training basins, the basin set is split into 100 random subsets of basins, with two-thirds used for calibration and one-third for validation. The functions shown in Table 1 are then fitted and subsequently validated on the basins not used for training. For each of the 100 trials, the R_{adj}^2 and the mean absolute error *MAE* between the true sensitivities derived using the xSSA analysis (see Fig. 3) and the predicted sensitivity based on the

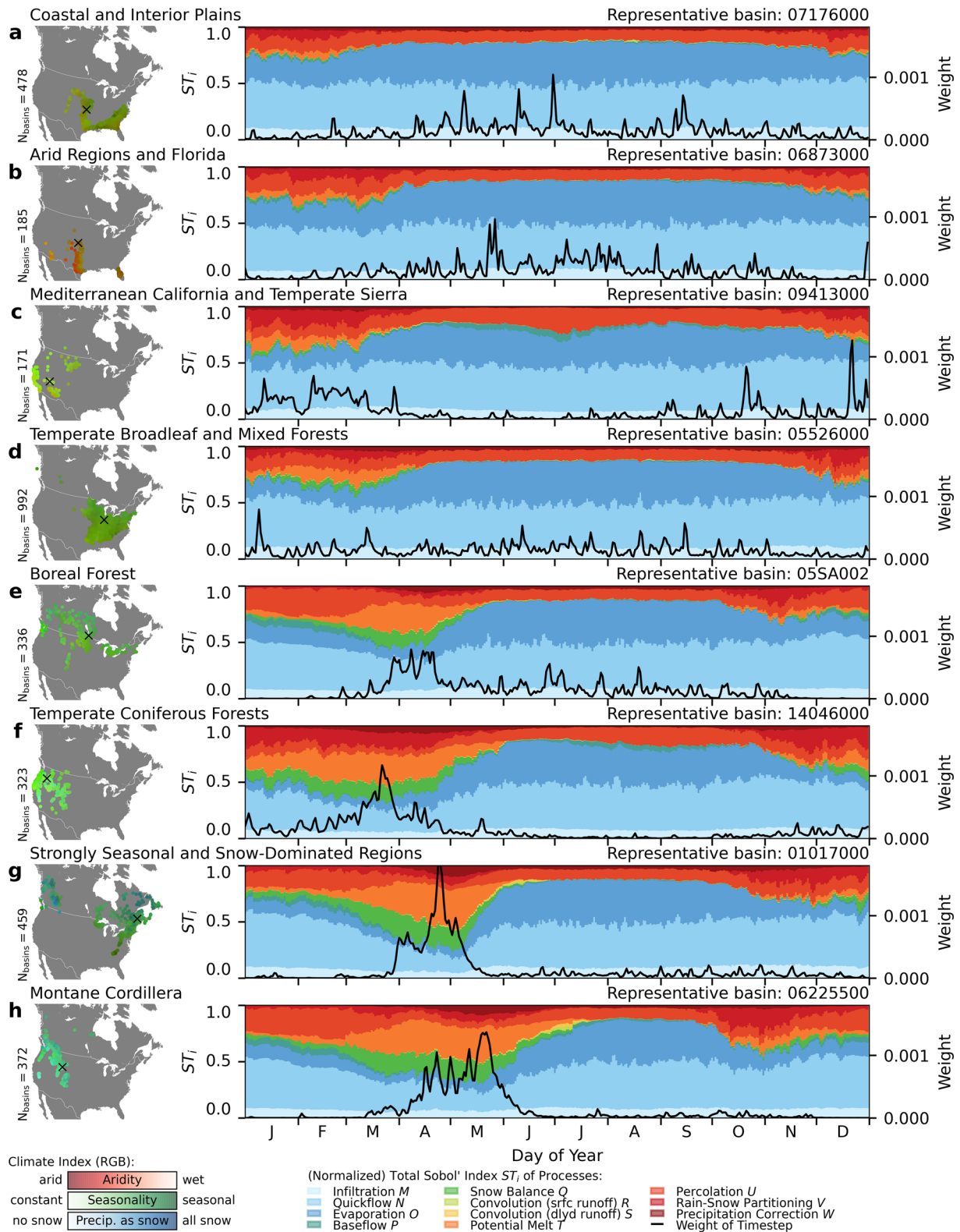


Fig. 4 Sensitivity of processes over time at selected locations. The importance of the eleven hydrologic processes (colorbars in right column panels) on each day of the year (average over 20 simulated years) are shown for eight example locations (a–h). These example locations are indicated by the USGS or WSC gauge station identifier added as a label at the top of the right column panels. These eight representative locations are the centers of eight clusters based on the three climate indicators. The clustering and example basins are derived to show the spatial variation of the process sensitivities over time in various climatic regions. Each cluster is named using the major physiographic or ecohydrologic region it covers (label on top of left column panels). For reference, the average weight of each day of the year based on the average simulated streamflow is displayed (black line in right column plots). These weights are used to derive time-aggregated sensitivities (see Fig. 3) from the time-dependent sensitivity indices displayed here in order to increase the weight of timesteps with high flows as compared to low-flow timesteps.

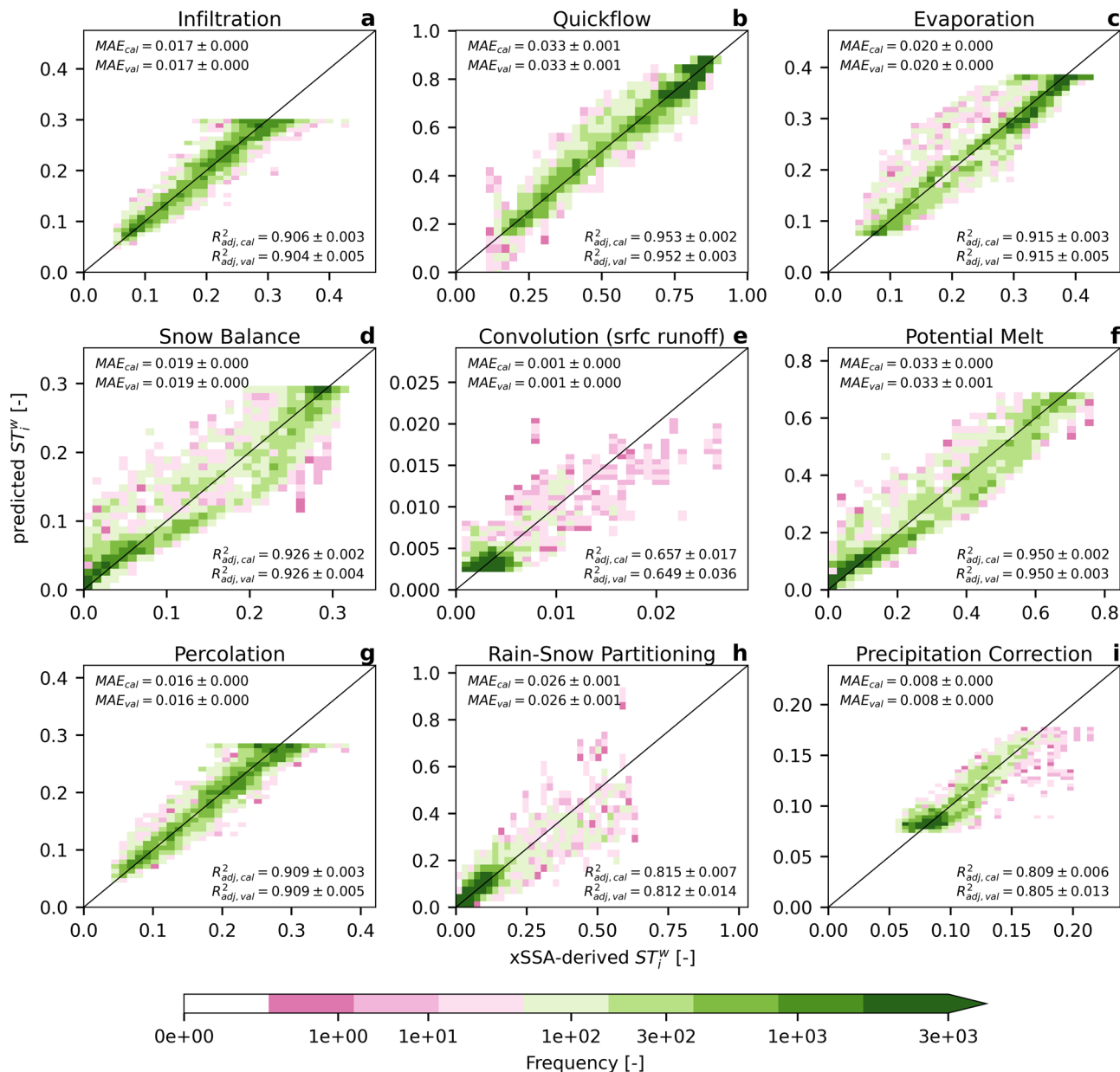


Fig. 5 Predicted sensitivity index estimates using functional relationships based on basin characteristics. The total Sobol' sensitivity indices of hydrologic processes on streamflow are first estimated using the xSSA method (Fig. 3; x-axis), after which they are predicted through basin-specific characteristics only (y-axis) and using the relations in Table 1. Each panel shows the results for one hydrologic process, i.e.: (a) infiltration, (b) quickflow, (c) evaporation, (d) snow balance, (e) convolution of surface runoff, (f) potential melt, (g) percolation, (h) rain-snow partitioning, and (i) precipitation correction. The mean and standard deviation of the Adjusted Coefficient of Determination R^2_{adj} and the mean absolute error (MAE) are reported. Both indexes are derived between (i) xSSA-derived indices and (ii) the predicted indices and added as labels for calibration (cal) and validation (val) sets. The color indicates the density of samples. The samples shown here are the samples of the 100 validation experiments.

regression are derived. The individual panels of the figure show the 2D histogram of 100 sets of validation basins (i.e., $100 \times 1/3 \times 3316 > 100,000$ data points). The two processes Baseflow and Convolution of delayed runoff are not analyzed as their sensitivity is lower than 0.01 for all basins.

All panels show that the predicted sensitivities are in close agreement with those derived by the xSSA analysis, indicating that just a handful of basic variables may be sufficient for describing the process sensitivity at most sites. The results are consistent between calibration cal and validation val for both fit metrics R^2_{adj} and MAE. Further, the 100 regression trials yield very consistent results, as indicated by the small standard deviations of the four metrics. The adjusted coefficients of determination R^2_{adj}

are above 0.80 for all processes in calibration and validation, except for Convolution of Surface Runoff (Fig. 5e; $R^2_{adj} \approx 0.65$), which is the least sensitive process. Only 88 of the 3316 basins show sensitivities larger than 0.01 for this process, which could explain the reduced overall predictability. In contrast, the very sensitive Infiltration (Fig. 5a), Quickflow (Fig. 5b), Evaporation (Fig. 5c), and Percolation (Fig. 5g) processes, in particular, show reliably large frequencies (dark green color) along the expected 1:1 line for large sensitivities. This, in turn, means that the processes with large sensitivity and importance can be identified and quantified reliably even without performing the expensive sensitivity analysis itself. The Potential melt (Fig. 5f), Precipitation correction (Fig. 5i), and Snow balance (Fig. 5d) processes,

Table 1 Deduced functional relationships using basin attributes to estimate process sensitivities.

Process	Pred. x	Pred. y	Estimated functional relationship $\widetilde{ST}_i^w = f(x, y)$	Adj. Coeff. of Determ. $R_{adj}^2(\widetilde{ST}_i^w, ST_i^w)$
Infiltration	f_s	-	$0.2985 - 0.8113x + 1.0443x^2 - 0.5203x^3$	0.905
Quickflow	f_s	ΣP	$0.9044 - 3.4122x^2 + 2.6596x^3 - 4.771 \times 10^{-5}y - 0.001153xy + 0.001870x^2y$	0.953
Evaporation	f_s	-	$0.3835 - 0.7713x + 0.1945x^2 + 0.3964x^3$	0.915
Snow Balance	f_s	-	$-0.009051 + 0.4161x + 1.3796x^2 - 2.0661x^3$	0.926
Convolution (srfc runoff)	\bar{T}	f_{cold}	$0.06531 - 0.00639x + 1.645 \times 10^{-4}x^2 - 0.0008165y + 4.104 \times 10^{-5}xy + 2.6508 \times 10^{-6}y^2$	0.657
Potential Melt	f_s	-	$-0.01494 + 0.8236x + 2.4265x^2 - 3.1523x^3$	0.950
Percolation	f_s	-	$0.2829 - 0.7275x + 0.8276x^2 - 0.3582x^3$	0.909
Rain-Snow Partitioning	f_s	f_{cold}	$-0.001715 + 3.7710x - 0.002150y - 0.03638xy + 1.302 \times 10^{-5}y^2 + 8.753 \times 10^{-5}xy^2$	0.815
Precipitation Correction	f_s	$I_{m,r}$	$0.07343 + 0.5829x^2 - 1.0106x^3 + 0.01122y^3 - 0.1677xy^2 + 0.5264x^2y$	0.809

Note. Results of the regression analysis identifying the functional relationships between predictors and the variance-weighted total Sobol' sensitivities ST_i^w based on the results of the 3316 basins analyzed by xSSA. The regression is performed using polynomials with one predictor (up to degree three with at most six coefficients) and polynomials with two predictors (up to degree three with at most five coefficients). The one-predictor polynomial is used unless the use of a pair of predictors led to an improvement of the adjusted coefficients of determination R_{adj}^2 by at least 0.05. The basin characteristics used as predictors are derived from geophysical attributes and the meteorology available for each basin (1950 to 2010). The adjusted coefficients of determination R_{adj}^2 between the xSSA-derived sensitivities ST_i^w and the predicted sensitivities \widetilde{ST}_i^w are given in the last column. Description of predictors: f_s is the fraction of precipitation that is snow [mm/mm], ΣP is the annual sum of precipitation [mm], \bar{T} is the annual average temperature in [°C], f_{cold} is the average annual number of days below 0 °C, $I_{m,r}$ is the measure of seasonality ranging between [0,2].

already identified to be important in the Rocky Mountains and the higher latitudes starting over the Great Lakes (see Fig. 3), are also reliably quantified, even for the rarer large sensitivity indices.

It is revealing to examine the most reliable sensitivity predictors from Table 1, which summarizes the best predictors identified by the regression analysis. The two processes not listed—Baseflow and Convolution of Delayed Runoff—have a total Sobol' sensitivity of 0.01 or less in each of the basins. The predictor with the largest influence is the fraction of precipitation as snow f_s , which is used as a predictor in eight of the nine relationships. This is in close agreement with the findings by Konapala et al.⁶⁴, who also identified the fraction of snow as a key characteristic to explain runoff signals in their study based on an information theory approach. The four processes that benefit from adding a second predictor are Quickflow, Convolution of Surface Runoff, Rain-Snow Partitioning and Precipitation Correction. While we do not want to read too much into the results for the almost insensitive Convolution of Surface Runoff process, the addition of the annual sum of precipitation ΣP to Quickflow, the average annual number of days below 0 °C f_{cold} to Rain-snow Partitioning, and seasonality $I_{m,r}$ to Precipitation Correction as a second predictor improved the R_{adj}^2 significantly (from 0.884 to 0.953, from 0.618 to 0.815, and from 0.650 to 0.809, respectively). These three secondary predictors are consistent with hydrologic reasoning.

In extreme cases, the regressions can result in negative sensitivity index estimates, which are known to be unrealistic. We suggest setting those values to zero in applications.

In summary, the deduced functional relationships between basin characteristics and the sensitivity of major hydrologic processes shows promising results, indicated by strong coefficients of determination R_{adj}^2 between derived and predicted sensitivities. The large number of basins, which can span over multiple climatic regions, ensures transferability of the relationships to North American basins on which an xSSA sensitivity analysis has not explicitly been performed. The consistency of results with previous studies indicates their robustness with respect to other models (e.g., PRMS) and sensitivity metrics (e.g., FAST method based on mean streamflow).

Discussion

This continental-scale sensitivity analysis of hydrologic model parameters and processes addresses the four key challenges raised in the introduction.

First, sensitivity analyses are traditionally carried out for model parameters only. This leads to difficulties in making high-level modeling decisions; these include, for example, whether to

prioritize model development and improvement or to secure better datasets, which may reduce uncertainty. The novel analysis used herein extends the traditional Sobol' sensitivity analysis to cover model components or processes at the continental scale.

Second, due to the computationally expensive nature of sensitivity analyses, they are usually carried out at a small set of locations. This work successfully applies the novel sensitivity analysis over a large domain, allowing us to draw conclusions across multiple climatic regions. The analysis leads to hydrologically consistent and quantitative sensitivity index estimates—either aggregated in time or time-dependent—identifying quickflow as the overall most sensitive process in the Eastern United States, with Infiltration, Evaporation and Percolation processes being of secondary importance. In snow-dominated regions such as the Rocky Mountains, the Potential melt, Snow balance, Precipitation correction, and Rain-snow partitioning processes are of large importance. The time-resolved sensitivity of processes provides particularly detailed insights into the common sensitivity of processes across all watersheds. Furthermore, the large number of basins analyzed allows to deduce functional relationships between basin characteristics and climatic indices that allow estimating the sensitivity of processes, even without the need to perform the computationally expensive sensitivity analysis for basins not covered herein.

Third, sensitivity analyses are usually based on a single model, which limits the conclusions to certain model process definitions and presents the risk that sensitivity estimates might be different for other models. The use of the Raven hydrologic modeling framework⁶⁵ and the novel blended model structure of⁴⁷ allow a seamless analysis of a range of model structures, thereby reducing the risk of inferring model-specific conclusions as a range of model structures are analyzed simultaneously. It should be noted that the model structure can be inferred during model calibration based on observations, to build hydrological modeling hypotheses (e.g., McMillan et al.⁶⁶, McMillan et al.⁶⁷, Clark et al.⁶⁸, Fenicia et al.⁶⁹, and Fenicia et al.⁷⁰). This work can help modelers decide on the model structure to use for inference of model structure in the following manner. First, processes with small sensitivities can be represented with essentially any option since the streamflow simulations are not affected in an important manner. Second, the processes with larger sensitivities should be included in the calibration process since the selection of the process option will play an important role in the streamflow simulations.

Fourth, sharing detailed results of large-scale model analyses is challenging due to the large amount and complexity of the data involved. The data and results used and produced in the work

presented herein are shared through an online mapping interface, which allows researchers to access and visualize inputs and results (see details under ‘Data Availability’). This includes all calibration setups and results, the climatic indicators and basin properties, as well as individual sensitivity results and summary plots.

The xSSA sensitivity method can be applied to any kind of model. A comparison against the only other continental-scale parameter-based sensitivity analysis we are aware of shows good consistency with our results. It would be of interest to investigate if more distributed land-surface hydrologic models were to show similar patterns in these process sensitivities; and further, if the process sensitivities could then also be related to basin characteristics, as shown in this study. It would additionally be interesting to investigate the influence on the sensitivities when other forcing datasets are used or other modeling decisions—such as different numbers of soil layers, other process options, or different simulation time steps—are made. Our approach to continental-scale sensitivity analysis can be emulated for more complex land-surface hydrologic models, and the results suggest such an analysis could be conducted over a reduced number of watersheds. Such a framework could be readily extended to other modular environmental and ecological models, similarly challenged by high degrees of structural uncertainty.

This study is a quantitative continental sensitivity analysis of streamflow simulations to hydrologic processes—transferable both to other (i) basins and (ii) models using similar process representations.

Methods

This work is out to analyze the sensitivity of streamflow to individual hydrologic processes across North America. This effort includes an attempt to deduce process sensitivities from basin characteristics alone without performing the expensive sensitivity analysis shown herein.

Blended model structure. The model used here is a blended model introduced by Mai et al.⁴⁷. The unique definition of the blended model within the Raven hydrologic modeling framework⁶⁵ enables the seamless simulation of various model structures simultaneously. A process, for example Infiltration, is no longer defined by one specific process algorithm (e.g., the infiltration definition used in model A), but is now calculated as the weighted average flux from several independent algorithms (for example, infiltration algorithms used by models A, B and C). The specific process options and associated model parameters used here are available in Tables C1 and C2 of Mai et al.⁴⁷, and are added to the Supplementary Material of this publication for the convenience of the reader. The blended model has 35 model parameters and eight parameters describing the weights between 19 process options overall. In total, the following eleven hydrologic processes are distinguished: Infiltration *M*, Quickflow *N*, Evaporation *O*, Baseflow *P*, Snow Balance *Q*, Surface Runoff *R*, Delayed Runoff *S*, Potential Melt *T*, Percolation *U*, Rain-Snow Partitioning *V*, and Precipitation Correction *W*. The model setup used in this study is exactly the same as in Mai et al.⁴⁷, except that here, it is applied to more than 3000 basins. More details about the model, its parameters and processes can be found in the Supplementary Material as well as in Mai et al.⁴⁷.

Basin database. The overall sequence of experiments performed in this study can be found in Fig. 1. The analysis is based on the HYSETS dataset⁴⁸, comprised of 14,425 basins filtered by size and overall data availability and quality. Details on the basin properties can be found on the website associated with this publication, showing eight physiographic attributes⁷¹ based on the void-filled HydroSHEDS digital elevation model (DEM) with a 30 arc-second resolution⁷², eight landcover types⁷³ based on the 250 m North American Land Change Monitoring System (NACLMS) for 2010⁷⁴, and three climatic indicators⁷⁵ used by Knoben et al.⁷⁶ for each basin.

Preliminary calibration and validation. A preliminary calibration/validation experiment of the blended model is performed to demonstrate the basic adequacy of the recently introduced blended model, and to act as the basis for excluding poorly performing models from analysis. This basic calibration using a single evaluation metric will be the precursor to a more elaborate future multi-metric model calibration/validation study to be informed by the results found here.

The subset of 3826 basins with more than five years of observed streamflow data between January 1991 and December 2010 (calibration period) is calibrated using the Dynamically Dimensioned Search (DDS)⁷⁷ algorithm maximizing the Nash-Sutcliffe efficiency (NSE)⁷⁸. A budget of 2000 model evaluations per basin and ten independent trials is used for DDS to calibrate the 35 model parameters and the 8 parameters defining the model structure. All model runs use a two-year warm-up

period (January 1989 to December 1990). Both calibration and sensitivity analysis results are found to be insensitive to warm-up period duration when this period is extended from 2 to 5 years (results not shown). The best performing parameter set of the ten trials is subsequently used for validation between January 1971 and December 1990. Only the 3005 basins with more than five years of observed streamflow available in this 20-year period are assessed in validation. The validation is performed to show that the overall performance during calibration is maintained during an independent time period. We assume that a strong validation performance indicates a low likelihood that the model parameters are overfitted during calibration. The general performance of the blended model in calibration and in validation compared to fixed model structures has been demonstrated in detail by Chlumsky et al.⁷⁹.

Sensitivity analysis. The 3316 basins with an NSE performance of 0.5 or higher during calibration are then analyzed regarding their simulated streamflow sensitivity. The method applied here is the extended Sobol’ Sensitivity Analysis (xSSA) introduced by Mai et al.⁴⁷, and illustrated at one example catchment per climate zone. This work replicates the analysis for 3316 basins, with exactly the same parameter and analysis settings. For each basin, $K = 1000$ Sobol’ reference sets were used, leading to a computational budget of $(43 + 2) \times K$ model runs to determine the sensitivities of the 43 parameters, $(27 + 2) \times K$ model runs to determine the sensitivities of the 19 process options as well as the eight parameters to weight these options, and $(11 + 2) \times K$ model runs to determine the overall sensitivity of the eleven processes. To assess the impact of a priori range specification for each parameter (Table S2 in Supplementary Material), we repeated the analysis for a subset of 150 basins using ranges for parameters x_2 to x_6 that were reduced by 20%.

The Sobol’ main sensitivity index S_i , as well as the total sensitivity index ST_i , are determined for the parameters, process options, and processes. Both indices are time-dependent since they are derived regarding the simulated streamflow time series $Q(t)$. To aggregate the indices over time, the variance-weighted approach as introduced by Cuntz et al.⁵⁹ and used by Mai et al.⁴⁷ is implemented. These indices are denoted by $S_i^w = \overline{S_i(Q(t))}$ and $ST_i^w = \overline{ST_i(Q(t))}$, respectively. Note that the averaging is performed over the sensitivity indices rather than the streamflow time series as used by Markstrom et al.³⁴, where the sensitivity regarding, for example, the mean runoff $\overline{S_i(Q(t))}$, is estimated. This averaging order is expected to lead to major differences as models and sensitivities are known to be non-linear.

The xSSA analysis is performed using the blended model setup within the Raven modeling framework⁶⁵, but can also be performed with any modeling framework, such as SUMMA^{80,81} and FUSE⁸². The process sensitivities can be estimated as soon as parameters are grouped into processes (no model implementation required), while properly analyzing multiple process options simultaneously requires the implementation of the weighted averages of these process option outputs (implementation in model most likely required).

Deduction of relationships between basin attributes and sensitivity analysis results.

Finally, we attempt to infer functional relationships between known basin characteristics and the importance of these hydrologic processes using regressions. The twelve basin characteristics used as predictors are the logarithmic area of the basin in [km^2], the mean elevation of the basin in [m], the average slope of the basin in [$^\circ$], the fraction of the forest cover in [km^2/km^2], the annual sum of precipitation in [mm], the annual average temperature in [$^\circ C$], the annual sum of potential evapotranspiration in [mm], the average annual number of days below $0^\circ C$, the average annual number of days where precipitation is sleet, i.e., temperature between $-0.85^\circ C$ and $1.15^\circ C$, as well as the aridity $I_{m,r}$, seasonality $I_{m,r}$, and fraction of precipitation as snow f_s . The latter three are derived following Knoben et al.⁷⁶. The regression analysis is performed using the Mathematica⁸³ function LinearModelFit by first fitting the 7 possible polynomials with up to degree three and up to five coefficients using each of the 12 combinations to pick one out of the twelve predictors. Second, we fit the 381 possible polynomials with up to degree three and up to six coefficients using each of the 66 combinations to pick two out of the twelve predictors. The regressions were performed in a cross-validation setup, i.e., each polynomial was fitted to two-thirds of the basins and validated using the remaining one-third of the basins. The polynomial with the best average adjusted coefficient of determination R_{adj}^2 in validation across 10 trials was identified— one best across all one-predictor polynomials and one best across the two-predictor polynomials. The latter two-predictor polynomial was chosen over the one-predictor version if the adjusted coefficient of determination was improved by at least 0.05. In total, 252,300 ($= 12 \times 7 \times 10 + 66 \times 381 \times 10$) regressions were performed per process predicting the process sensitivity ST_i^w of the nine processes. The two Baseflow and Convolution of delayed runoff processes are excluded as their weighted total Sobol’ index ST_i^w are smaller than 0.01 in all of the 3316 basins analyzed. The regression function leading to the best adjusted coefficient of determination R_{adj}^2 is then used to perform a regression using all 3316 basins (reported in Table 1). To assess the robustness of this regression, we performed another cross-validation experiment similar to the previous one. The regression coefficients of the best regression (reported in Table 1) are refit now using two-thirds of the basins and validating them using the remaining one-third of the basins (reported in Fig. 5). This experiment is repeated with 100 random splits of basins.

Data availability

The xSSA data generated in this study, including examples, can be found on GitHub (<https://github.com/julemai/xSSA-North-America>). They have been deposited in the Zenodo database under accession code <https://doi.org/10.5281/zenodo.5730428>. Additional information such as model setups and interactive visualization of data and results can be found on the webpage associated with this publication (http://www.hydrohub.org/sa_introduction.html#xssa-na) and in the Supplementary Material.

Code availability

The xSSA code used to generate the results presented in this study, including examples, can be found on GitHub (<https://github.com/julemai/xSSA-North-America>). They have been deposited in the Zenodo database under accession code <https://doi.org/10.5281/zenodo.5730428>.

Received: 20 May 2021; Accepted: 3 January 2022;

Published online: 24 January 2022

References

- Weerts, A. H. & El Serafy, G. Y. H. Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resour. Res.* **42**, W09403 (2006).
- Todini, E. A model conditional processor to assess predictive uncertainty in flood forecasting. *Int. J. River Basin Manag.* **6**, 123 (2008).
- Bogner, K. & Pappenberger, F. Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. *Water Resour. Res.* **47**, W07524 (2011).
- Yucel, I., Onen, A., Yilmaz, K. & Gochis, D. Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction model, data assimilation and satellite-based rainfall. *J. Hydrol.* **523**, 49 (2015).
- Rakovec, O., Weerts, A. H., Sumihar, J. & Uijlenhoet, R. Operational aspects of asynchronous filtering for flood forecasting. *Hydrol. Earth Syst. Sci.* **19**, 2911 (2015).
- Rogelis, M. C. & Werner, M. Streamflow forecasts from WRF precipitation for flood early warning in mountain tropical areas. *Hydrol. Earth Syst. Sci.* **22**, 853 (2018).
- Svoboda, M. et al. The Drought Monitor. *Bull. Am. Meteorol. Soc.* **83**, 1181 (2002).
- Zink, M. et al. The German drought monitor. *Environ. Res. Lett.* **11**, 074002 (2016).
- Aadhar, S. & Mishra, V. Data descriptor: high-resolution near real-time drought monitoring in South Asia. *Sci. Data* **4**, 1 (2017).
- Zink M. et al. A National Scale Planning Tool for Agricultural Droughts in Germany, in *Advanced Tools for Integrated Water Resources Management*. Vol. 3, 147–169 (Elsevier, 2018).
- Dietrich, A., Yarlagadda, R. & Gruden, C. Estimating the potential benefits of green stormwater infrastructure on developed sites using hydrologic model simulation. *Sust. Energ.* **36**, 557 (2017).
- Avellaneda, P. M., Jefferson, A. J., Grieser, J. M. & Bush, S. A. Simulation of the cumulative hydrological response to green infrastructure. *Water Resour. Res.* **53**, 3087 (2017).
- Fry, T. J. & Maxwell, R. M. Using a distributed hydrologic model to improve the green infrastructure parameterization used in a lumped model. *Water* **10**, 1 (2018).
- Yang, S.-C. & Yang, T.-H. Uncertainty assessment: reservoir inflow forecasting with ensemble precipitation forecasts and HEC-HMS. *Adv. Meteorol.* **2014**, 1 (2014).
- Schwabenberg, D. et al. Short-term reservoir optimization for flood mitigation under meteorological and hydrological forecast uncertainty. *Water Resour. Manag.* **29**, 1635 (2015).
- Mai, J., Arsenault, R., Tolson, B. A., Latraverse, M. & Demeester, K. Application of parameter screening to derive optimal initial state adjustments for streamflow forecasting. *Water Resour. Res.* **56**, 1 (2020a).
- Schoups, G., van de Giesen, N. C. & Savenije, H. H. G. Model complexity control for hydrologic prediction. *Water Resour. Res.* **44**, W00B03 (2008).
- Orth, R., Staudinger, M., Seneviratne, S. I., Seibert, J. & Zappa, M. Does model performance improve with complexity? A case study with three hydrological models. *J. Hydrol.* **523**, 147 (2015).
- Li, H., Xu, C. Y. & Beldring, S. How much can we gain with increasing model complexity with the same model concepts? *J. Hydrol.* **527**, 858 (2015).
- Coron, L., Andréassian, V., Perrin, C., Bourqui, M. & Hendrickx, F. On the lack of robustness of hydrologic models regarding water balance simulation: a diagnostic approach applied to three models of increasing complexity on 20 mountainous catchments. *Hydrol. Earth Syst. Sci.* **18**, 727 (2014).
- Clark, M. P. et al. The evolution of process-based hydrologic models: historical challenges and the collective quest for physical realism. *Hydrol. Earth Syst. Sci.* **21**, 3427 (2017).
- Menard, C. B. et al. Scientific and human errors in a snow model intercomparison. *Bull. Am. Meteorol. Soc.*, 10.1175/BAMS-D-19-0329.1 (2020).
- Mai, J. et al. Great Lakes runoff intercomparison project phase 3: Lake Erie (GRIP-E). *J. Hydrol. Eng.*, **26**, 05021020 (2021).
- Gupta, H. V., Wagener, T. & Liu, Y. Reconciling theory with observations: elements of a diagnostic approach to model evaluation. *Hydrol. Process.* **22**, 3802 (2008).
- Hrachowitz, M. et al. Process consistency in models: the importance of system signatures, expert knowledge, and process complexity. *Water Resour. Res.* **50**, 7445 (2014).
- Nijzink, R. et al. Constraining conceptual hydrological models with multiple information sources. *Water Resour. Res.* **54**, 8332 (2018).
- Khatami, S., Peel, M. C., Peterson, T. J. & Western, A. W. Equifinality and flux mapping: a new approach to model evaluation and process representation under uncertainty. *Water Resour. Res.* **55**, 8922 (2019).
- Széles, B. et al. The added value of different data types for calibrating and testing a hydrologic model in a small catchment. *Water Resour. Res.* **56**, e2019WR026153 (2020).
- Dembéle, M. et al. Potential of satellite and reanalysis evaporation datasets for hydrological modelling under various model calibration strategies. *Adv. Water Resour.* **143**, 103667 (2020).
- Mendoza, P. A. et al. Are we unnecessarily constraining the agility of complex process-based models? *Water Resour. Res.* **51**, 716 (2015).
- Demirel, M. C. et al. Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. *Hydrol. Earth Syst. Sci.* **22**, 1299 (2018).
- Göhler, M., Mai, J. & Cuntz, M. Use of eigendecomposition in a parameter sensitivity analysis of the community land model. *J. Geophys. Res. Biogeosci.* **118**, 904 (2013).
- Cuntz, M. et al. The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model. *J. Geophys. Res. Atmos.* **121**, 10676–10700 (2016).
- Markstrom, S. L., Hay, L. E. & Clark, M. P. Towards simplification of hydrologic modeling: identification of dominant processes. *Hydrol. Earth Syst. Sci.* **20**, 4655 (2016).
- Saltelli, A. et al. *Global Sensitivity Analysis: The Primer* (Wiley-Interscience, 2008).
- Ferretti, F., Saltelli, A. & Tarantola, S. Trends in sensitivity analysis practice in the last decade. *Sci. Total Environ.* **568**, 666 (2016).
- Saltelli, A. et al. Why so many published sensitivity analyses are false: a systematic review of sensitivity analysis practices. *Environ. Model. Softw.* **114**, 29 (2019).
- Razavi, S. & Gupta, H. V. What do we mean by sensitivity analysis? The need for comprehensive characterization of ‘Global’ sensitivity in Earth and Environmental Systems Models. *Water Resour. Res.* **51**, 3070–3092 (2015).
- Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Zeng, X. & de Gonçalves, L. G. G. A fully multiple-criteria implementation of the Sobol’ method for parameter sensitivity analysis. *J. Geophys. Res. Atmos.* **117**, 1 (2012).
- Herman, J. D., Kollat, J. B., Reed, P. M. & Wagener, T. From maps to movies: high-resolution time-varying sensitivity analysis for spatially distributed watershed models. *Hydrol. Earth Syst. Sci.* **17**, 5109 (2013).
- Haghnegahdar, A., Razavi, S., Yassin, F. & Wheeler, H. Multicriteria sensitivity analysis as a diagnostic tool for understanding model behaviour and characterizing model uncertainty. *Hydrol. Process.* **31**, 4462 (2017).
- Schürz, C. et al. A comprehensive sensitivity and uncertainty analysis for discharge and nitrate-nitrogen loads involving multiple discrete model inputs under future changing conditions. *Hydrol. Earth Syst. Sci.* **23**, 1211–1244 (2019).
- Rakovec, O., Hill, M. C. & Clark, M. P. Distributed evaluation of local sensitivity analysis (DELSA), with application to hydrologic models. *Water Resour. Res.* **50**, 409 (2014).
- Chard, K. et al. Globus Data Publication as a Service: Lowering Barriers to Reproducible Science, in *2015 IEEE 11th International Conference on eScience* 401–410 (IEEE, 2015).
- Vasilevsky, N. A., Minnier, J., Haendel, M. & Champieux, R. E. Reproducible and reusable research: are journal data sharing policies meeting the mark? *PeerJ* **5**, e3208 (2017).
- Persaud, B. D. et al. Ten Best Practices to Strengthen Stewardship and Sharing of Water Science Data in Canada. *Hydrol. Process.* **35**, 14385 (2021).
- Mai, J., Craig, J. R. & Tolson, B. A. Simultaneously determining global sensitivities of model parameters and model structure. *Hydrol. Earth Syst. Sci.* **24**, 5835 (2020b).
- Arsenault, R. et al. A comprehensive, multisource database for hydrometeorological modeling of 14,425 North American watersheds. *Sci. Data* **7**, 1 (2020).
- Rakovec, O. et al. Diagnostic evaluation of large-domain hydrologic models calibrated across the contiguous United States. *J. Geophys. Res. Atmos.* **124**, 13991 (2019).

50. Samaniego, L., Kumar, R. & Attinger, S. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resour. Res.* **46**, W05523 (2010).
51. Kumar, R., Samaniego, L. & Attinger, S. Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. *Water Resour. Res.* **49**, 360 (2013).
52. Mizukami, N. et al. Towards seamless large-domain parameter estimation for hydrologic models. *Water Resour. Res.* **53**, 8020 (2017).
53. Liang, X., Lettenmaier, D. P., Wood, E. F. & Burges, S. J. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *J. Geophys. Res. Atmos.* **99**, 14415 (1994).
54. Liang, X. A new parameterization for surface and groundwater interactions and its impact on water budgets with the variable infiltration capacity (VIC) land surface model. *J. Geophys. Res.* **108**, 1989 (2003).
55. Newman, A. J. et al. 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. *Hydrol. Earth Syst. Sci.* **19**, 209 (2015).
56. Burnash, R. J. C. & Singh, V. *Computer Models of Watershed Hydrology* (Water Resources Publication, 1995).
57. Anderson, E. A. *Calibration of Conceptual Hydrologic Models for Use in River Forecasting*. NOAA Technical Report NWS 45. (Hydrology Laboratory, Silver Spring, 2002).
58. Juliane Mai, xSSA for North America: Calibration results, http://www.hydrohub.org/xssa-na/maps_calibration_map.html (2021a), accessed: November 26, 2021.
59. Cuntz, M. et al. Computationally inexpensive identification of noninformative model parameters by sequential screening. *Water Resour. Res.* **51**, 6417 (2015).
60. Markstrom, S. L. et al. PRMS-IV, the Precipitation-Runoff Modeling System, Version 4., in *U.S. Geological Survey Techniques and Methods, book 6, chap. B7* (U.S. Department of the Interior, U.S. Geological Survey, 2015) p. 158.
61. Juliane Mai, xSSA for North America: Aggregated total Sobol' Sensitivity indexes, http://www.hydrohub.org/xssa-na/maps_xSSA_aggregated_map.html (2021b), accessed: November 26, 2021.
62. Bezdek, J. C. Pattern recognition with fuzzy objective function algorithms, Advanced applications in pattern recognition (Plenum Press, 1981).
63. Juliane Mai, xSSA for North America: Time-dependent total Sobol' Sensitivity indexes, http://www.hydrohub.org/xssa-na/maps_xSSA_time-dependent_map.html (2021c), accessed: November 26, 2021.
64. Konapala, G., Kao, S.-C. & Addor, N. Exploring hydrologic model process connectivity at the continental scale through an information theory approach. *Water Resour. Res.* **56**, 1 (2020).
65. Craig, J. R. et al. Flexible watershed simulation with the Raven hydrological modelling framework. *Environmen. Model. Softw.* **129**, 104728 (2020).
66. McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M. & Woods, R. A. Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure. *Hydrol. Process.* **25**, 511 (2011).
67. McMillan, H. et al. Spatial variability of hydrological processes and model structure diagnostics in a 50 km² catchment. *Hydrol. Process.* **28**, 4896 (2014).
68. Clark, M. P., McMillan, H. K., Collins, D. B. G., Kavetski, D. & Woods, R. A. Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses. *Hydrol. Process.* **25**, 523 (2011).
69. Fenicia, F. et al. Catchment properties, function, and conceptual model representation: is there a correspondence? *Hydrol. Process.* **28**, 2451 (2014).
70. Fenicia, F., Kavetski, D., Savenije, H. H. G. & Pfister, L. From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions. *Water Resour. Res.* **52**, 954 (2016).
71. Juliane Mai, xSSA for North America: Basin properties: Physiographic, http://www.hydrohub.org/xssa-na/maps_basin-property_physiographic_map.html (2021d), accessed: November 26, 2021.
72. Lehner, B., Verdin, K. & Jarvis, A. New Global Hydrography Derived From Spaceborne Elevation Data. *EOS Trans. Am. Geophys. Union* **89**, 93 (2008).
73. Juliane Mai, xSSA for North America: Basin properties: Landcover, http://www.hydrohub.org/xssa-na/maps_basin-property_landcover_map.html (2021e), accessed: November 26, 2021.
74. CEC CCA CCE, Land cover, 2005 (modis, 250m) (2010), accessed: 2021-03-03 <https://doi.org/10.5067/MODIS/MCD12Q1.006>
75. Juliane Mai, xSSA for North America: Basin properties: Knoben climate indicators, http://www.hydrohub.org/xssa-na/maps_basin-property_knoben-climate-indicator_map.html (2021f), accessed: November 26, 2021.
76. Knoben, W. J. M., Woods, R. A. & Freer, J. E. A quantitative hydrological climate classification evaluated with independent streamflow data. *Water Resour. Res.* **54**, 5088 (2018).
77. Tolson, B. A. & Shoemaker, C. A. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resour. Res.* **43**, W01413 (2007).
78. Nash, J. E. & Sutcliffe, J. V. River flow forecasting through conceptual models: Part I - A discussion of principles. *J. Hydrol.* **10**, 282 (1970).
79. Chlumsky, R., Mai, J., Craig, J. R. & Tolson, B. A. Simultaneous calibration of hydrologic model structure and parameters using a blended model. *Water Resour. Res.* **57**, e2020WR029229 (2021).
80. Clark, M. P. et al. A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resour. Res.* **51**, 2498 (2015a).
81. Clark, M. P. et al. A unified approach for process-based hydrologic modeling: 2. Model implementation and case studies. *Water Resour. Res.* **51**, 2515 (2015b).
82. Clark, M. P. et al. Framework for understanding structural errors (FUSE): a modular framework to diagnose differences between hydrological models. *Water Resour. Res.* **44**, 2135 (2008).
83. W. R. Inc., Mathematica, Version 12.1.1.0 (2020).

Acknowledgements

This research was undertaken thanks in part to funding from the CANARIE research software funding program (project RS-332). The work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET; www.sharcnet.ca) and Compute/Calcul Canada. The authors also thank the Canadian Foundation for Innovation John Evans Leaders Fund for the additional supercomputing support and resources. We thank Pawel Pomoski for the Dedicated Programming Support granted through the "SHARCNET Round XVI Programming Competition" to improve the performance of the xSSA code implementation and the addition of parallel computing capabilities.

Author contributions

J.M. set up the models, implemented and performed model calibrations and sensitivity analyses, wrote main parts of the manuscript, prepared all figures and tables, created the website hosting all setups and results; J.R.C. contributed to the writing of the manuscript, implemented required modifications in Raven, provided feedback on model setups, helped to set up the model with the selected options and resolved inconsistencies in Raven detected by earlier versions of the sensitivity analysis, and helped with the hydrologic interpretation of the results; B.A.T. contributed to the experimental design, the writing of the manuscript, provided feedback on the manuscript and helped with the hydrologic interpretation of the results; R.A. provided meteorologic forcings and geophysical data to set up the watersheds used in this study and helped with the hydrologic interpretation of the results.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-28010-7>.

Correspondence and requests for materials should be addressed to Juliane Mai.

Peer review information *Nature Communications* thanks Rajtantra Lilhare and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022