RESEARCH ARTICLE

Cumulative Effects of Uncertainty on Simulated Streamflow in a Hydrologic Modeling Environment

Scott Pokorny*, Tricia A. Stadnyk†, Genevieve Ali‡, Rajtantra Lilhare§, Stephen J. Déry§‖ and Kristina Koenig¶

It is common in the literature to not consider all sources of uncertainty simultaneously: input, structural, parameter, and observed calibration data uncertainty, particularly in data-sparse environments due to data limitations and the complexities that arise from data limitations when propagating uncertainty downstream in a modelling chain. This paper presents results for the propagation of multiple sources of uncertainty towards the estimation of streamflow uncertainty in a data-sparse environment. Uncertainty sources are separated to ensure low likelihood uncertainty distribution tails are not rejected to examine the interaction of sources of uncertainty. Three daily resolution hydrologic models (HYPE, WATFLOOD, and HEC-HMS), forced with three precipitation ensemble realizations, generated from five gridded climate datasets, for the 1981–2010 period were used to examine the effects of cumulative propagation of uncertainty in the Lower Nelson River Basin as part of the BaySys project. Selected behavioral models produced an average range of Kling-Gupta Efficiency scores of 0.79–0.68. Two alternative methods for behavioral model selection were also considered that ingest streamflow uncertainty. Structural and parameter uncertainty was found to be insufficient, individually, by producing some uncertainty envelopes narrower than observed streamflow uncertainty. Combined structural and parameter uncertainty, propagated to simulated streamflow, often enveloped nearly 100% of observed streamflow values, however, high and low flow years were generally a source for lower reliabilities in simulated results. Including all sources of uncertainty generated simulated uncertainty bounds that enveloped most of the observed flow uncertainty bounds including improvement for high and low flow years across all gauges although the uncertainty bounds generated were of low likelihood. Overall, accounting for each source of uncertainty added value to the simulated uncertainty bounds when compared to hydrometric uncertainty; the inclusion of hydrometric uncertainty was key for identifying the improvements to simulated ensembles.

Keywords: Uncertainty; Uncertainty Propagation; Hydrologic modeling; Hydrology; BaySys

1. Introduction

Hydrologic models are used to generate many different types of output, most frequently streamflow. Models make various simplifications when representing a target physical environment (e.g. Clark et al., 2015), and these simplifications, among the other sources, introduce uncertainty to simulated output. Uncertainty is defined as the realistic range to which an exact value for a given variable cannot be determined but can be represented by a likelihood (Uusitalo et al., 2015). Uncertainty is generated by a model’s structure (e.g., Clark et al., 2015) and its parameters (e.g., Shafii et al., 2015). Uncertainty is also ingested through model input (e.g., Pokorny, 2019) and observational data used for calibration, hereafter referred to as hydrometric uncertainty (e.g., McMillan et al., 2010). Uncertainties interact within a hydrologic modeling framework by propagation. Propagation is the process of how uncertainty from one modeling step affects the next, moving cumulatively downstream (e.g., Brown and Heuvelink, 2006; Ajami et al., 2007; Nikolopoulos et al., 2010; Mei et al., 2016). All methods for quantifying uncertainty are limited by subjectivity (Kavetski et al., 2003; Kavetski et al., 2006; Beven and Binley, 2014). Examination of the subjectivity in uncertainty estimation methods is a common topic that has driven the development of the many uncertainty estimation frameworks available in the literature (Matott et al., 2009).
Input uncertainty includes a wide range of data sources. Many studies choose to focus on precipitation as a dominant source of input data uncertainty (e.g., Kavetski et al., 2006; Ajami et al., 2007). Some studies consider meteorological data errors and uncertainty towards suggesting a preferred product (e.g., Choi et al., 2009; Eum et al., 2014, Wong et al., 2017), while others suggest meteorological input ensembles (e.g., Montanari and Di Baldassarre, 2013; Rapaï et al., 2015; Gambia et al., 2017; Lilhare et al., 2019). Meteorologic input ensembles are preferred as they better account for uncertainty than a single realization (Kavetski et al., 2006; Ajami et al., 2007). Meteorologic ensembles are common in hydrologic climate change studies (e.g., Dams et al., 2015; Karlsson et al., 2016), but less so in historical hydrologic studies (e.g., Nikolopoulos et al., 2010).

Model structural uncertainty represents the uncertainty generated by model spatial and temporal aggregation (Dams et al., 2015; Wi et al., 2015; Muhammad et al., 2018a), hydrologic process numerical representation (Tasdighi et al., 2018), and hydrologic connectivity, which considers internal model decision-making for the allocation of water and the order of calculation (e.g., Clark et al., 2015). Structural uncertainty is often accounted for by multi-model studies (e.g., Ajami et al., 2007), or modular modeling (Clark et al., 2015).

Parameter uncertainty is inherently tied to structural uncertainty because the number and influence of parameters are dictated by the numerical process representation used in a hydrological model. Parameter uncertainty is often presented with parameter likelihood distributions described by the Generalized Likelihood Uncertainty Estimation (GLUE) methodology (Beven and Binley, 1992). The GLUE methodology has been the subject of much discussion due to the use of both formal and informal likelihood functions (Beven and Binley, 2014). A statistical inference may be ill-posed due to miss-classifications of uncertainty, or limited representation of uncertainty (Beven, 2016). The limits of acceptability (LOA; Beven, 2006) methodology address some of these shortcomings; however, subjectivity remains (e.g., Li et al., 2010; Li and Xu, 2014; Shafii et al., 2015), which may be further complicated by parameter identifiability (Wagener et al., 2003; Abebe et al., 2010; Merz et al., 2011). An informal likelihood function with GLUE may still offer important insights when assumptions for formal inference are not satisfied (Beven and Binley, 2014). The informal objective function and model selection method will influence the selected parameter sets (e.g., Shafii et al., 2015).

Hydrometric uncertainty, like input uncertainty, is ingested by a hydrologic model. Hydrometric uncertainty is a function of multiple sources of uncertainty such as gauge accuracy, rating curve fit, and many others (Shiklomanov et al., 2006; Hamilton, 2008; Hamilton and Moore, 2012; McMillan et al., 2012; Westerberg et al., 2016; Whitfield and Pomeroy, 2017; McMillan et al., 2018). Methods to quantify uncertainty in hydrometric data exist and are still being developed in recent literature and focus on rating curve based uncertainty estimates (e.g., Coxon et al., 2015; Kiang et al., 2018). Hydrometric uncertainty is not accounted for in traditional calibration metrics (e.g., Gupta et al., 2009; Westerberg et al., 2020). The LOA methodology can include hydrometric uncertainty but relies on subjective decision thresholds for model rejection. Reliability and sharpness offer an alternative objective function to optimize for simulation acceptance and rejection (Li et al., 2010; Westerberg et al., 2011; Li et al., 2014; Bourgin et al., 2015; Shafii et al., 2015; Zhou et al., 2016). Reliability is a measure of overlap with an observed benchmark (Montanari, 2005), and sharpness is defined as a measure of uncertainty bound width (Yadav et al., 2007).

Studies often find input or structural uncertainty to be the largest contributors to total uncertainty (e.g., Ajami et al., 2007; Chen et al., 2011; Dams et al., 2015). No study has exhaustively sampled all possible sources of uncertainty due to dimensionality issues, data limitations, and computational limitations (Ajami et al., 2007). Therefore, total uncertainty is defined as the accumulation of all sources of uncertainty considered in a study; this definition is used in this study as well. Separating the contribution of each source of uncertainty is difficult due to the complex interactions among sources; sampling techniques are generally applied to consider each source of uncertainty after propagation to output (Pechlivanidis et al., 2011). For instance, models like WATFLOOD (Kouwen, 2018) use grouped response units, an approach that ties parameter values (parameter uncertainty) to landcover (input uncertainty) (e.g. Eckhardt et al., 2003). Pokorny (2019) shows that precipitation data uncertainty (input uncertainty) is tied to spatial aggregation, which is also a function of model structure (Wi et al., 2015; Muhammad et al., 2018a). Bayesian inference methods are often applied to generate uncertainty distributions with formal likelihood functions (Stedinger et al., 2008; Renard et al., 2010). Formal methods, however, often present poorly constrained problems arising from various epistemic sources of uncertainty being simplified to stationary aleatory errors (Beven, 2016). Model output is also subjected to the propagation of all sources of uncertainty, which suggests uncertainty distributions are relative rather than absolute (Beven and Binley, 2014). The information gained by generating relative uncertainty distributions still informs of relative importance for the sampling of a particular source of uncertainty (e.g. Kavetski et al., 2003; Kavetski et al., 2006; Renard et al., 2010; Dams et al., 2015). Indeed, the relative separation of uncertainty sources has been widely studied in the literature (e.g. Kavetski et al., 2003; Vrugt et al., 2005; Kavetski et al., 2006; Clark et al., 2011; Clark et al., 2015). Comparisons of uncertainty sources and their cumulative effects on propagated uncertainty as they compound to total modeling uncertainty, however, remain rarely studied.

Few studies examine the cumulative impacts of each of the sources of uncertainty as they propagate to model output (e.g. Ajami et al., 2007). This is especially true in large, northern basins in Canada where data sparsity compounds the difficulties associated with the generation of relative uncertainty distributions. The consideration of low likelihood uncertainty distribution tails is generally reserved for forecasting based studies, but, they also become more important when data sparsity is an issue.
(e.g. Demeritt et al., 2007; Cloke and Pappenberger, 2009; Addor et al., 2011; Pappenberger et al., 2013). For example, rating curve uncertainty estimates for Canadian hydrometric data are not currently available due to data limitations (Hamilton and Moore, 2012). The Water Survey of Canada suggests ±5% uncertainty (Environment Canada, 1980), however, the literature suggests more complexity in uncertainty estimates up to ±10% (Dingman, 2015) or higher (e.g. McMillan et al., 2010; McMillan et al., 2012). Wide bounds of uncertainty are often criticized for lacking value because the likelihood of those bounds is low, hence communication of the value of wide uncertainty bounds is a common issue (Pappenberger et al., 2013).

To address the aforementioned gaps in the literature, the objectives of this study are to 1) select a general method appropriate for characterizing uncertainty propagation in remote data-sparse regions, 2) apply this method to generate relative uncertainty distributions for multiple sources of hydrologic model uncertainty, and 3) examine the cumulative effects of propagated sources of uncertainty on simulated flow uncertainty including low likelihood uncertainty distribution tails. This study is part of the Hudson Bay System (BaySys) project (see Barber, 2014). The BaySys project is a multi-disciplinary regulations impact study in which the effects of climate change and hydropower regulation are relatively partitioned for comparison. To generate large scale relative partitions of hydropower regulation and climate change impacts, the results from numerical models (e.g. hydrologic models) are used as inputs to other numerical models (e.g. ocean models). With model output being ingested by other models as input, understanding the uncertainty generated by those models is important for determining the value of final results. The region of study used to address the study objectives is the Lower Nelson River Basin (LNRB), which is further described in the following sections. The LNRB is a highly relevant basin to the BaySys project, as it is subject to considerable regulation and suffers from data sparsity.

2. Study area
The region of study is the Lower Nelson River Basin (LNRB) in Northern Manitoba, Canada (Figure 1). The LNRB was selected for its BaySys project relevance and because of its relative data sparsity (Barber, 2014); beyond the BaySys consideration, any data-sparse region could have been considered. The LNRB is a basin of ~90,500 km$^2$ at the

![Figure 1: The Lower Nelson River Basin (blue shading) with available hydrometric stations with historical data available, including the locations of generating stations, and the currently under construction Keeyask generating station. The locations of Thompson and Gillam are highlighted by arrows and the period average nudged flow magnitudes added to the models are shown in text near the nudging locations. Nudged inflow from Jenpeg and the East Channel are combined. DOI: https://doi.org/10.1525/elementa.431.f1](https://doi.org/10.1525/elementa.431.f1)
downstream end of the ~1,400,000 km² Nelson Churchill Watershed (NCW) (Figure 1). The LNRB is characterized by a sub-arctic climate and low-relief terrain largely covered by forest and wetlands, with many lakes due to the low topographic relief. The LNRB is a data-sparse basin with limited climate stations. January is the coldest month with 30-year average temperatures (1981–2010) of −23.9°C and −24.4°C in the towns of Thompson and Gillam, respectively. Summers are cool, with July being the warmest month (1981–2010) with temperatures of 16.2°C and 15.8°C at Thompson and Gillam, respectively. Precipitation (1981–2010) is highest in July at 80.9 mm and 78.6 mm for Thompson and Gillam, respectively. Total annual precipitation for Thompson and Gillam is 509.0 mm and 496.4 mm, respectively, for the 1981–2010 period, of which 43% occurred in summer (JJA). Hydrograph peak flows are dominated by spring melt runoff. Spring peaks in unregulated basins generally occur in May except for the Grass River, in which peak flows are usually attenuated to July or later by lakes and wetlands. The LNRB, and upstream contributing area, have been included in numerous climate change studies, which suggest hydrologic vulnerability to climate change. Some notable studies looked at shifts in river ice jams (e.g., Rokaya et al., 2018), changes in water volume originating from the prairies (e.g., Muhammad et al., 2018a), hydrologic regime trends (e.g., Westmacott et al., 1997), flood peaks (e.g., Clavet-Gaumont et al., 2017), among others. For an in-depth review of the climatology and physiography of the LNRB, see Smith (2015), Holmes (2016), and Lilhare et al. (2019).

3. Data and models

3.1. Hydrometric data

Hydrometric gauges are shown in Figure 1 and detailed in Table 1. Water from Lake Winnipeg enters the LNRB via the Nelson River through the Jenpeg Generating Station and the East Channel and is augmented by flow diverted from the Churchill River (Notigi control structure) via the Burntwood River. There are six Manitoba Hydro operated generating stations within the LNRB: the Jenpeg, Kelsey, Kettle, Long Spruce, Limestone, and Wuskwatim generating stations acting as points of regulation. Additionally, there is the Keeyask Generating Station, which is currently under construction. Manitoba Hydro operates the Wuskwatim generating station on behalf of the Wuskwatim Power Limited Partnership, of which Manitoba Hydro is a partner, and will operate the Keeyask generating station as a partner. Publicly available WSC data were supplemented with internal Manitoba Hydro records. Missing data were infilled using linear interpolation for short gaps of two or fewer days. Longer data gaps were not infilled; performance metrics were calculated based on available data.

Since the LNRB is a downstream basin in the NCW, the flow from upstream sources must be added to the basin and considered as an input. The flow records from the Jenpeg generating station, East Channel, and the Churchill River Division (Notigi Control Structure) are, therefore, added as model forcings to the LNRB hydrologic models. The flows estimated through Jenpeg and the Churchill River Division are measured reasonably accurately, control structures provide stable conditions for flow measurement. The inflows from Jenpeg and the Churchill diversion are closely monitored and regulated, which means measurement uncertainty is small. Control structures are fixed with known dimensions, which allows uncertainty to approach gauge accuracy. With such ideal streamflow estimation conditions, uncertainty can reasonably be assumed negligible. The East Channel is subjected to the same uncertainty of any other non-regulated gauge, however, it represents only a small portion of the forced inflows. Therefore, while nudged inflows are subject to minor uncertainty, it was assumed negligible, which also reduced computational time for model runs.

3.2. Gridded climate data ensemble

Meteorological input data for this study consists of daily timeseries of precipitation and air temperature (daily; 1981–2010). The ensemble consists of five gridded data-sets, presented in Table 2, which are suggested to be reasonable representations of observed conditions in the literature (Lilhare et al., 2019; Pokorny, 2019).

All gridded datasets were bilinearly interpolated to match ANUSPLIN’s 0.10° grid (Lilhare et al., 2019). Daily time steps were used when available. For NARR, daily minimum and maximum air temperatures were estimated from 3-hourly data. Three precipitation realizations were generated: Ensemble minimum, mean, and maximum. Minimum (maximum) precipitation was generated by selecting the minimum (maximum) value of all five grid-ded datasets for each grid cell at each time-step (i.e., daily). This method is expected to generate an overestimate of uncertainty. Ideally, an estimate of total uncertainty would be available for every gridded meteorological dataset but this is not computationally feasible. Recent studies suggest that the selected datasets suffer from notable uncertainty (Lilhare et al., 2019). Therefore, since the goal is to ensure low likelihood uncertainty distribution tails are not rejected, the wide range of precipitation uncertainty is assumed reasonable. A mean precipitation realization was generated by the arithmetic mean of all five datasets with equal weighting. Regarding other model inputs, only the arithmetic mean air temperature realization was considered, which reduced computational demands, as uncertainty was noticeably lower in the estimation of temperatures, relative to the uncertainty in precipitation (Pokorny, 2019). Uncertainty data associated with temperature input are not considered because temperature data is measured with less uncertainty than precipitation. It is a standard assumption that precipitation data uncertainty is the major contributor to meteorological input uncertainty, which is built into most uncertainty frameworks (e.g., Kavetski et al., 2006; Huard and Mailhot, 2006; Ajami et al., 2007; Vrugt et al., 2008; Renard et al., 2010; McMillan et al., 2011).

3.3. Hydrologic model ensemble

An ensemble of hydrologic models, all running on daily timesteps, was selected to account for structural uncertainty by varying both spatial aggregation and hydrologic
Table 1: Hydrometric gauges in the LNRB considered in this study. All data are considered at a daily timestep. DOI: https://doi.org/10.1525/elementa.431.t1

| Station Name                                      | ID      | Longitude (°W) | Latitude (°N) | Gauged Drainage Area (km²) | Mean Annual Flow (m³ s⁻¹) | Regulated |
|--------------------------------------------------|---------|----------------|----------------|-----------------------------|---------------------------|-----------|
| **Hydrometric Gauges**                           |         |                |                |                             |                           |           |
| 1) Footprint River Above Footprint Lake¹         | 05TF002 | 98.88          | 55.93          | 643                         | 3                         | No        |
| 2) Taylor River Near Thompson                    | 05TG002 | 98.19          | 55.48          | 886                         | 5                         | No        |
| 3) Kettle River Near Gillam                      | 05UF004 | 94.69          | 56.34          | 1090                        | 13                        | No        |
| 4) Angling River Near Bird                       | 05UH001 | 93.64          | 56.68          | 1560                        | 11                        | No        |
| 5) Weir River Above The Mouth                    | 05UH002 | 93.45          | 57.02          | 2190                        | 16                        | No        |
| 6) Limestone River Near Bird                     | 05UG001 | 94.21          | 56.51          | 3270                        | 22                        | No        |
| 7) Bumtwood River Above Leaf Rapids              | 05TE002 | 99.22          | 55.49          | 5810                        | 23                        | No        |
| 8) Odie River Near Thompson                      | 05TG003 | 97.35          | 55.99          | 6110                        | 34                        | No        |
| 9) Grass River above Standing Stone Falls        | 05TD001 | 97.01          | 55.74          | 15400                       | 65                        | No        |
| 10) Bumtwood River Near Thompson                 | 05TG001 | 97.90          | 55.74          | 18500                       | 867                       | Yes       |
| 11) Nelson River at Kelsey Generating Station    | 05UE005 | 96.59          | 55.94          | 1050000                     | 2350                      | Yes       |
| 12) Nelson River at Kettle Generating Station¹   | 05UF006 | 94.37          | 56.40          | 1100000                     | 3550                      | Yes       |
| 13) Nelson River at Long Spruce Generating Station | 05UF007 | 94.37          | 56.40          | 1100000                     | 3550                      | Yes       |
| 14) Rat River Below Notigi Control Structure     | 05TF003 | 99.29          | 55.86          | 6140                        | 790                       | Yes       |
| 15) Nelson River (West Channel) at Jenpeg        | 05UB009 | 98.05          | 54.50          | 974500                      | 1880                      | Yes       |
| 16) Nelson River (East Channel) Below Sea River Falls² | 05UB008 | 97.59          | 54.24          | 976000                      | 361                       | Yes       |

¹These stations were not used due to close proximity to other stations, therefore, they did not provide additional information.
²Station 05UB008 is a natural channel but is considered regulated by the WSC because its water source (Lake Winnipeg) is subject to regulation. Therefore, it is affected by regulation but not regulated itself.
process representation. The models selected to represent the LNRB included: the Hydrological Predictions for the Environment (HYPE; SMHI, 2018), the WATFLOOD model (Kouwen, 2018), and the Hydrologic Engineering Center – Hydrologic Modeling System (HEC-HMS; USACE, 2016) (Table 1). Existing models of the LNRB were used with some modifications to parameterizations. Differences in underlying data used to set up the models, such as land cover, are assumed to be minor and are accounted for with re-parameterizations (Eckhardt et al., 2003). The assumption that non-meteorological input selection (and differences) among the model setups introduce minor uncertainties has more of an impact for WATFLOOD and HYPE, since parameters in these models are tied to land and soil classification, but is less impactful to HEC-HMS. Assuming negligible temporal changes in land and soil classification is of concern when performing long-term simulation (e.g., 50 or 100-years), but has less of an impact over a shorter duration climatological period. Regardless, detailed land and soil classification for the 1981–2010 period were not available for the LNRB (Dwarakish and Ganasri, 2015). Contributing areas upstream of hydrometric gauges were reviewed and adjusted to ensure equal comparisons between models.

The HYPE and WATFLOOD models were set up for long periods of simulation (i.e. 30 years) (Stadnyk et al., Accepted; Holmes, 2016). The HYPE model was set up as part of the BaySys project for historic and future simulations. The WATFLOOD model was set up for isotope-based calibration. However, the version of the LNRB HEC-HMS model used was set up for flood studies. Therefore, the HEC-HMS hydrologic processes were updated to make use of the Priestly-Taylor evapotranspiration method (Priestley and Taylor 1972), which is expected to perform better than the fixed monthly evapotranspiration method used by Sagan (2017) for the 30 year simulation period used in this study (Table 3). The hydrologic models were run at a daily timestep for the 1981–2010 climatic period. Two additional years of model spinup were included (1979–1980).

| Dataset | Description | Reference | Temporal resolutions; Available period; Spatial resolution | Domain |
|---------|-------------|-----------|----------------------------------------------------------|---------|
| ANUSPLIN | The Australian National University spline interpolation | Hutchinson et al., 2009 | Daily; 1950–2011; 0.10° | Canada – Land |
| NARR | North American Regional Reanalysis | Mesinger et al., 2006 | 3-hourly and Daily; 1979-near present; 0.30° | North America |
| HydroGFD | Hydrological Global Forcing Data | Berg et al., 2018 | 3-hourly, 6-hourly, and Daily; 1979-near present; 0.50° | Global – Land |
| WFDEI | European Union Water and Global Change (WATCH) Forcing Data ERA-Interim | Weedon et al., 2014 | 3-hourly, 6-hourly, and Daily; 1979-near present; 0.50° | Global |
| ERA-Interim | European Centre for Medium-Range Weather Forecasts interim reanalysis | Dee et al., 2011 | 3-hourly, 6-hourly, and Daily; 1979-near present; 0.75° | Global |

Table 3: Summary of key structural differences between the selected hydrologic models. Studies describing the model setups are included in parentheses. DOI: https://doi.org/10.1525/elementa.431.13

| Hydrologic Model | Hydrologic Processes | Selected methods | Comments |
|------------------|----------------------|-----------------|----------|
| HYPE (Stadnyk et al., Accepted) | Infiltration | HYPE default infiltration | Semi-lumped sub-basin model with basin sizes generally around 400 km² |
| | Evapotranspiration | Priestly-Taylor | |
| | Snowmelt | Temperature + Radiation Index | |
| | Routing | Lag, Recession, and Attenuation | |
| WATFLOOD (Holmes, 2016) | Infiltration | Phillips Formula | Gridded model with 10 km grid spacing |
| | Evapotranspiration | Hargreaves | |
| | Snowmelt | Temperature Index | |
| | Routing | Storage routing | |
| HEC-HMS (Sagan, 2017) | Infiltration | Soil Moisture Accounting | Semi-lumped sub-basin model with basin sizes ranging from 360–12,000 km² |
| | Evapotranspiration | Priestly-Taylor | |
| | Snowmelt | Temperature Index | |
4. Methodology

The methodology for this study is designed to maintain low likelihood uncertainty distribution tails. Separate consideration of each source of uncertainty was conducted to ensure lower likelihood simulations were not rejected.

4.1. Data preparation

Input datasets include the three precipitation realizations, the arithmetic mean temperature realization, and the unaltered observed streamflow records at nudging locations. HEC-HMS required daily dewpoint temperatures for the Priestly-Taylor method. Dewpoint temperatures were not available in all datasets; a comparable mean product could not be generated for dewpoint temperatures. Therefore, dewpoint temperatures were generated from precipitation and temperature data so that the effect of using the minimum and maximum precipitation realizations was reflected in the estimate of dewpoint temperatures. Dewpoint temperature timeseries were estimated using Equation (1), developed by Hubbard et al. (2003):

\[ T_d = \alpha (T_a) + \beta (T_s - T_a) + \gamma (P_{\text{daily}}) + \lambda, \]  

where \( T_d \) was the daily dewpoint temperature (°C); \( T_a \) was the daily minimum temperature (°C); \( T_s \) was the daily maximum temperature (°C); \( P_{\text{daily}} \) was the daily precipitation (mm); and \( \alpha, \beta, \gamma, \) and \( \lambda \) were unitless regression constants fitted to ground-based climate station dewpoint temperature data in the LNRB by least squares regression.

Input data for the HYPE model, recently set up in part by Stadnyk et al. (Accepted), was generated by assigning the nearest climate data grid points to sub-basins. The input data for the WATFLOOD LNRB model, recently set up in part by Holmes (2016), aligned with the 0.10° gridded ensemble data. The HEC-HMS LNRB model used input climate data generated by aggregating grid points within and along basin delineations.

4.2. Uncertainty analysis

Input uncertainty was addressed by the three precipitation input realizations; precipitation was considered the focus as its uncertainty is assumed higher than other model inputs (Wong et al., 2017). The precipitation ensemble minimum and ensemble maximum represent higher uncertainty than would be useful for design or operational work. The wide uncertainty bounds generated by the ensemble minimum and the ensemble maximum represent low likelihood tails of the uncertainty distribution. The low likelihood distribution tails may increase in likelihood under climate change effects, however, the uncertain range was also more similar to the uncertainty generated in medium-range ensemble flood forecasting (e.g., Han and Coulibaly, 2019). Different percentile ranges can be used to consider narrower ranges of uncertainty while retaining the ability to consider low likelihood events at higher percentile ranges (e.g. Pappenberger et al., 2013).

Uncertainty introduced by model structure was addressed through the application of three hydrologic models of varied structures. For example, the Odei River basin was represented by multiple sub-basins in HYPE (semi-lumped), by 82 0.10° grid cells, including those downstream of the streamflow gauge, by WATFLOOD (semi-distributed), and by a single basin in HEC-HMS (lumped). The range of spatial structures among the models tested the effect of spatial aggregation on input uncertainty propagation. Additionally, each model used different methods to represent individual hydrologic processes, each having a different number and type of parameters (Table 1). This model structure sampling does not allow for detailed partitioning of structural uncertainty introduced separately from spatial and temporal aggregation, selection of numerical processes, and hydrologic connectivity (internal flow paths). Additional sampling would have been required, varying a single of these sources with the others held constant, which is a sampling technique better suited to modular modeling (Clark et al., 2015).

Instead, structural uncertainty is considered through the selection of three different model structures, which varies each type of structural uncertainty (aggregation, process, and connectivity) simultaneously; generating a single estimate of the overall structural uncertainty from the ensemble of models. This presents a potential limitation: using three established model structures does not ensure structural uncertainty extends to low likelihood distribution tails.

Model parameters are used to tune model performance to be representative of a target physical environment; however, they are subject to equifinality (Beven and Binley, 1992). Parameter uncertainty was addressed using the GLUE methodology (Beven and Binley, 1992; Beven and Binley, 2014); parameter sets were sampled using orthogonal Latin-hypercube sampling (OLHS; Tang 1993) from uniform priors bounded by ranges suggested by the most recent modelers (Holmes, 2016; Sagan, 2017; Stadnyk et al., Accepted). A total of 23, 51, and 63 parameters were selected for HYPE, WATFLOOD, and HEC-HMS, respectively. Sample quantities were determined, in part, by sensitivity analyses (Holmes, 2016; Sagan, 2017; Stadnyk et al., Accepted), and by computational resources.

Some parameters were grouped by land cover or soil characteristics to further reduce computational demand. A total of 6900, 15,300, and 18,900 parameter samples were run for HYPE, WATFLOOD, and HEC-HMS, respectively (i.e. 2300 samples for each of three precipitation inputs with HYPE). More samples were given to models with more parameters because the response surface becomes more complex as the number of parameters increases. These samples were held constant for each precipitation input. A simple top 10% criterion applied to Kling-Gupta Efficiency (KGE) scores is used to select behavioral parameter sets (Gupta et al., 2009); the criterion, however, is subjective, which is a commonly debated topic in the literature (e.g., Steedinger et al., 2008; Li et al., 2010; Li and Xu, 2014; Shafii et al., 2015). An advantage to the top 10% criterion is the generation of wider uncertainty bounds than selection by optimization of an ensemble metric (Shafii et al., 2015). The KGE metric...
such as rating curves or error estimates, are not available but is limited by hydrometric data availability. Metadata, Moore, 2012; McMillan et al., 2012; Kiang et al., 2018),

which, 

\[ w(t) = \begin{cases} 
\frac{Q_{\text{sim}}(t)}{Q_{\text{obs}}(t)} & \text{if } Q_{\text{sim}}(t) < Q_{\text{obs}}(t) \\
1 & \text{if } Q_{\text{obs}}(t) \leq Q_{\text{sim}}(t) \leq Q_{\text{obs}}(t) \\
\frac{Q_{\text{obs}}(t)}{Q_{\text{sim}}(t)} & \text{if } Q_{\text{sim}}(t) > Q_{\text{obs}}(t)
\end{cases} \tag{3}
\]

which, \( w(t) \) and \( Q_{\text{obs}}(t) \) are weights and observed streamflow values for timestep \( t \) of \( T \) total timesteps, respectively. Weights are determined by Equation 3, where \( Q_{\text{sim}}(t) \) is the simulated streamflow for timestep \( t \), and \( Q_{\text{obs}}(t) \) and \( Q_{\text{obs}}(t) \) are the lower and upper bounds of hydrometric uncertainty, respectively.

Integrating hydrometric uncertainty into an objective function does not reduce the subjectivity of behavioral simulation selection; the top 10% of simulations is still selected. A second selection criterion is applied here as the combined overlap percentage (\( \text{COP} \), Equation 4) presented by Westerberg et al. (2011):

\[ \text{COP} = \frac{\sum_{t=1}^{T} \text{mean} \left( \frac{Q_{\text{obs}}}{Q_{\text{obs}}} - \frac{Q_{\text{obs}}}{Q_{\text{obs}}} \right)}{T} \tag{4} \]

which, \( Q_{R} \) is the overlap with the observed hydrometric uncertainty range \( Q_{\text{obs}} \), and \( Q_{\text{sim}} \) is the simulated uncertainty bound range. Simulations are selected that maximize the \( \text{COP} \) metric, which reduces the likelihood range. Formal likelihood functions were not considered to avoid simplifying epistemic errors to narrow the parameter uncertainty range.

The estimation method for observed hydrometric data uncertainty was adopted from McMillan et al. (2012) and simplified for this study. Open water flows and ice-affected flows were assumed to be determined by the “B” flag (Environment and Climate Change Canada 2019). Open water hydrometric flow data were assigned ±10% uncertainty bounds, whereas winter ice-affected flows were assigned ±20% uncertainty bounds. The estimate of observed streamflow uncertainty is an oversimplification (Hamilton 2008; McMillan et al., 2010; Hamilton and Moore, 2012; McMillan et al., 2012; Kiang et al., 2018), but is limited by hydrometric data availability. Metadata, which are defined as additional relevant information such as rating curves or error estimates, are not available for Canadian hydrometric data at this time. Total uncertainty was assessed using reliability and sharpness metrics (e.g., Montanari 2005; Yadav et al., 2007; Gneiting et al., 2007; Shafii et al., 2015; Zhou et al., 2016). Reliability was defined as the percentage of overlap of the observed flow bounds (daily) and the simulated ensemble bounds (daily) for the full period (1981–2010). Sharpness was defined in this study as (Equation 5):

\[ \text{Sharpness} = \frac{1}{T} \sum_{t=1}^{T} \frac{Q_{\text{obsUpper}} - Q_{\text{obsLower}}}{Q_{\text{simUpper}} - Q_{\text{simLower}}}, \tag{5} \]

in which \( Q_{\text{simUpper}} \) and \( Q_{\text{simLower}} \) are the upper and lower bounds of the simulated ensemble, respectively, and \( Q_{\text{obsUpper}} \) and \( Q_{\text{obsLower}} \) are the upper and lower observed flow uncertainty bounds. Reliability and sharpness values are calculated using the minimum and maximum ensemble bounds at each timestep, although some studies used a percentile range. A sharpness value of one means simulated uncertainty has, on average, the same uncertainty range as observed uncertainty. A sharpness value less than one indicates that simulated uncertainty exceeded observed uncertainty. Theoretically, the only way for both reliability and sharpness, as defined here, to approach one over the long simulation period of 30 years would be for a hydrologic model ensemble to approach perfection in its representation of the physical environment, which is not feasible.

In practice, short term periods of adverse measurement conditions elevate hydrometric uncertainty (e.g., ice on conditions, river ice breakup, overbank flooding, beaver dams, vegetation growth, channel morphology, etc.). This increased hydrometric uncertainty will not necessarily be reproduced by the hydrologic model, meaning a sharpness value above one is not always bad. With the simple estimate of hydrometric uncertainty implemented in this study, hydrometric uncertainty is likely under- or over-represented at each timestep, but becomes a reasonable average to represent long-term (1981–2010) hydrometric uncertainty. Still, short term events like rating curve extrapolation (Kiang et al., 2018) or ice jams (Hamilton and Moore, 2012) generate periods of very high hydrometric uncertainty. The simple estimate of hydrometric uncertainty applied in this study will not capture such short term events; we, therefore, present sharpness values for the full 1981–2010 study period to reduce the impact of such events in our analysis.

If a simulated ensemble generates high-reliability values but very low sharpness values, the uncertainty range extends into the low uncertainty tails of the likelihood uncertainty distribution (Beven and Binley, 2014); simulations at the uncertainty distribution tails are often excluded as they are of sufficiently low likelihood to not be valuable for design or general operations. Comparisons of the cumulative effect of uncertainty were assessed with reliability and sharpness plots partitioned by binned 10 percentile ranges; for example, flows between the 10th and 20th percentiles were grouped and represented at the 15th percentile. Total uncertainty was presented for an average of the five highest flow volume years and for an average of the five lowest flow volume years.
5. Results

5.1. Performance of behavioral simulations
The performance varied for each model (Table 4); HEC-HMS has the largest range in KGE for five locations, WATFLOOD has the largest range in KGE values for four locations, and HYPE has the largest range in KGE values for two locations. This result is anticipated and is a reflection of the number of parameters considered for each model. Regulated gauges (shown in bold) generally have the highest KGE scores and have a low range of KGE values since flows from upstream contributing areas are added as model input at nudging locations. Only additional flow generated downstream of those locations (typically smaller, proportionately) could affect regulated gauge performance. Simulations selected with the $F_{\text{obj}}$ metric generally produced negative scores close to zero, meaning simulated output in a given timestep was often outside the range of hydrometric uncertainty, but reasonably represented observed streamflow data. COP scores for each model were generally above 0.4.

5.2. Structural and parameter uncertainty
Reliability and sharpness (Figure 2) for each model are representative of parameter uncertainty propagated to model output, for each model. Structural uncertainty propagated to simulated output is represented by selecting the highest performing simulation for each model. The structural and parameter-based ensemble is representative of the combination of propagated parameter and structural uncertainty. Results suggested three general uncertainty source relationships, reliability dominated by a single model, reliability dominated by different models for different flow percentiles, and complimentary performance, in which no single model had high reliability, but ensemble reliabilities were still high. An example of each relationship is presented in Figure 2 as well as one regulated gauge.

The Taylor River reliability plot is generally dominated by a single model; the WATFLOOD model has similar reliability to that of the ensemble up to the 70th percentile flows. The Taylor River sharpness plot reflects that...
WATFLOOD's high reliability is due to wide-ranging model performance, supported by Table 2, in which WATFLOOD exhibits the widest range of KGE values. HEC-HMS has the lowest performance for the Taylor River, and a similar but slightly smaller sharpness range. The low performance of HEC-HMS is sourced from a consistent wet bias; no HEC-HMS simulations have a negative bias, and the lowest bias is 4.2% in contrast to WATFLOOD, which has simulations with both positive and negative biases. The Taylor River sharpness plot for HYPE shows that the HYPE model alone is not able to represent the full uncertainty range of the observed data, although this is generally the sharpest gauge for HEC-HMS and WATFLOOD as well since the Taylor River is the smallest basin, and is dominated by forest, which simplified the parameter uncertainty for that gauge.

The Weir River reliability plot (Figure 2) indicates that multiple models are of value to the ensemble reliability; WATFLOOD is better at representing observed uncertainty for flows below the 60th percentile, while HEC-HMS is better at representing flows above the 60th percentile. As reliability increases, sharpness generally decreases.

The Burntwood above Leaf Rapids location shows no model to be of high reliability between the 40th and 70th percentiles; however, the ensemble reliability is still close to 100%. Different model structures are complementary to the ensemble reliability; however, the decrease in sharpness of the ensemble versus any individual model is also generally high. For flows below the 40th percentile, HYPE and WATFLOOD are of high reliability but are not the models with the lowest sharpness. Unlike the high reliability of WATFLOOD in the Weir River, which is sourced from wide uncertainty bounds, WATFLOOD's high reliability in the Burntwood above Leaf Rapids region is sourced from narrow uncertainty bounds, relative to HEC-HMS and HYPE.

For the regulated Nelson at Long Spruce gauge, the models are complementary to the ensemble's reliability. Sharpness values for the Nelson at Long Spruce show the models generally produced uncertainty bounds that are narrower than the observed uncertainty bounds. At this location, however, observed flows are added for contributing areas upstream of the LNRB at nudging points, and that uncertainty from these upstream locations is not propagated downstream or reflected in the model results.

Propagated structural uncertainty always has higher reliability than at least one of the individual model's parameter uncertainty. For the Burntwood above Leaf Rapids and Nelson at Long Spruce Generating Station gauges (Figure 2), structural uncertainty is more valuable to the combined ensemble than any model's parameter uncertainty, for three and seven of the selected percentile bins, respectively. The average reliability across all stations and percentiles for structural uncertainty is 48%; as is the average reliability for parameter uncertainty, averaged across all three models. The average sharpness across all gauges and percentiles for structural and parameter uncertainty are 1.8 and 2.4, respectively.

Simulations selected by the F_{obj} metric generated slightly lower reliabilities for low percentiles, but slightly higher reliabilities for high percentiles when compared to the KGE-selected simulations (Appendix A; Figure A1). Since the F_{obj} metric gives higher weight to high flows, improved higher flow reliability was expected. Simulations generally produced sharper uncertainty bounds than simulations selected using the KGE. Simulations selected by optimized COP were generally sharper than both the KGE and F_{obj} simulations without notable loss of reliability (Appendix A; Figure A2). All ensembles selected using the COP consisted of fewer simulations than the number of simulations selected by the top 10% selection criterion (e.g. <230 simulations for HYPE). Simulations were not selected
based on individual performance with the COP, so no best performing simulation for each model could be selected to assess structural uncertainty. Since the goal was to ensure low likelihood uncertainty distribution bounds were not rejected, generating a structural uncertainty estimate does not limit the interpretation of the COP selected simulations.

5.3. Input uncertainty

The uncertainty generated from varying model structures and parameters propagated to streamflow varies in magnitude (Figure 2 and Table 2). When the wide range of input uncertainty is included, reliabilities generally increase to near 100% for unregulated gauges across all three models for all simulation selection criteria. The percentile ranges of sharpness with input data uncertainty are, on average: 8.8, 6.3, 3.0 and 4.1 times lower for HYPE, WATFLOOD, HEC-HMS, and the structural and parameter ensemble, respectively for the KGE selected simulations; sharpness was reduced for all simulation selection methods. These changes are relative and do not mean HYPE has the widest uncertainty bounds with input uncertainty included. With reliability values at or near 100% and low sharpness, the uncertain range extends to the tails of the likelihood distribution. Reliabilities for each of the three regulated gauges are generally >60% for all flow percentiles and simulation selection methods. Regulated gauge reliabilities are not as affected by input uncertainty as are non-regulated gauges, because the regulated gauges are affected by flow nudging. Results for the Weir River with simulations selected by KGE are presented. The Weir River was selected since it displayed some of the three general uncertainty relationships highlighted by parameter and structural reliability and sharpness metrics (Figure 3). The width of the flow envelope in each panel is a representation of uncertainty propagated through a hydrologic model.

HEC-HMS has the widest parameter uncertainty range in more basins than WATFLOOD or HYPE (e.g., Figure 3j versus Figure 3f). Input uncertainty is widest for WATFLOOD and narrowest for HEC-HMS, which is generally consistent across all gauges and simulation selections, although there was some variability. The range of structural uncertainty is generally narrower for the minimum precipitation input, and wider for the maximum precipitation input. Areas with zero density (i.e., gaps between hydrographs) are the result of only sampling at the extreme ranges of input uncertainty (i.e., minimum, mean, and maximum); if further sampling is done, the uncertainty range (i.e. reliability) would not change, but the areas of zero density would fill in.

5.4. Total modeling uncertainty

On average, most of the hydrometric uncertainty bounds are enveloped by the 10th and 90th percentiles of the combined structural and parametric uncertainty (Figure 4a and b). Hydrometric flow bounds that lie outside the 10th...
and 90th percentiles are generally below the simulated ensemble for low flow volume years (Figure 4a) and above the simulated ensemble for high flow volume years (Figure 4b); this is consistent for all simulation selection methods. Including input uncertainty notably widens the simulated uncertainty bounds for both low flow and high flow volume years (Figure 4c and d). Most of the low flow observed bounds are bracketed by the 25th and 75th percentiles of the total simulated ensemble, however, the 10th and 90th percentiles are required to envelop the spring runoff peaks in the high flow volume years. The total simulated ensemble generally has the highest simulated density near the observed flow bounds, while much lower density is present near the uncertainty bounds.

The COP selection criterion generates a much sharper total uncertainty estimate (Figure 5). Simulations that introduce timing issues were rejected by the COP criterion. COP values for total uncertainty were generally above 0.6, but many of the low likelihood uncertainty distribution tails were rejected. Similar to the KGE and \( F_{\text{obj}} \) selected simulations, high (low) flow years were often under- (over-) estimated.

6. Discussion
6.1. Cumulative effects of uncertainty propagation
Models selected from the top 10%, rather than selected by optimization generated uncertainty bounds that are wider than when optimized by the COP metric (e.g., Stedinger et al., 2008; Shafii and Tolson 2015). Narrow uncertainty bounds are considered desirable (Shafii and Tolson, 2015), however, without consideration of hydrometric uncertainty, bounds may be too narrow to adequately represent hydrometric data uncertainty. The combination of parameter and structural uncertainty (Figure 2) is an improvement over parameter uncertainty alone for reliability for all simulation selection criteria (Ajami et al., 2007; Chen et al., 2011; Muhammad et al., 2018b; Muhammad et al., 2019). An important consideration is, however, the quality of the hydrometric uncertainty estimate. The simple estimate for hydrometric uncertainty is oversimplified but is reasonable when considered over the 30-year period, and preferable to not considering hydrometric uncertainty at all. Each hydrologic model structure has temporally variant strengths that, when considered as an ensemble (e.g. Figure 2), improve reliability (Ajami et al., 2007). The propagated structural uncertainty generally has reliability higher than at least one parameter uncertainty estimate and is generally one of the sharpest uncertainty partitions, which suggests that structural uncertainty would be identified as more important than parameter uncertainty if hydrometric uncertainty were not considered (e.g. Ajami et al., 2007, Chen et al., 2011, and Dams et al., 2015).

Precipitation ensemble inputs present wide uncertainty bounds. When propagated through the hydrologic
models, the resulting simulated uncertainty bounds approach 100% reliability for all simulation selection criteria. It is common to consider a wide percentile range such as the 95% prediction range (Shafii and Tolson, 2015), however, narrower percentile ranges are also presented in the literature, such as 35th and 65th percentile bounds presented by Han and Coulibaly (2019). The choice of which percentiles to use is based on communication goals (Pappenberger et al., 2013). Wide uncertainty bounds include low likelihood events that are of importance (particularly when considering long-term climate change), but if presented improperly, may lower confidence in the simulated output (Demeritt et al., 2007). The 10th, 25th, 75th, and 90th percentiles are presented in Figure 4c and d. The 10th and 90th represent a wide range of uncertainty for most of the year but are needed to bracket some of the highest and lowest flow events. The simulated range of uncertainty was found to be similar to that of some flood forecasting uncertainties (Han and Coulibaly, 2019), or those considering alternative meteorological data sources, like satellite-derived precipitation inputs (e.g., Nikolopoulos et al., 2010).

Therefore, while the addition of input uncertainty generates a wide range of uncertainty, the improvement in reliability suggests that all sources of uncertainty should be considered to better represent hydrometric streamflow uncertainty. In data-sparse regions, there is often not enough information to constrain uncertainty. Low-density observational networks adversely impact calibration for all environmental models (e.g. Hutchinson et al., 2009). Constraining uncertainty to produce sharp uncertainty bounds is almost always the goal of studies considering uncertainty (e.g. Ajami et al., 2007), however, for flood forecasting, wide uncertainty bounds are generated and uncertainty is constrained by considering narrower percentile ranges in post-processing steps (Han and Coulibaly, 2019). In forecasting studies, forecasted meteorologic conditions present a large source of (input) uncertainty that remains challenging to constrain. Narrow percentiles like 35th and 65th may be considered if there is little consequence for misrepresentation of a flow event (usually at specified quantiles of interest). If there are notable consequences for misrepresenting a given quantile event that occurs at low likelihoods, wider percentiles are used instead. If low likelihood simulations are rejected, then wider percentiles cannot be considered (e.g. Figure 5).

Connecting this to the BaySys project, in addition to a remote, data-sparse study region, the goal was to produce a wide range of uncertainty that could be considered across various percentile ranges to explore the potential impact of freshwater extremes on the marine system circulation (Barber, 2014). As more high-quality data are made available through remote sensing techniques, more of the lower likelihood uncertainty distribution tails could be rejected from the selection criteria. In data-sparse regions, however, consideration of low likelihood simulations adds value to interpretations of results. This result would also apply to other data-sparse regions; however, for areas with higher data density, lower likelihood events could perhaps

Figure 5: Uncertainty bounds compared with estimated hydrometric data uncertainty bounds for the Weir River (average daily). Results are presented for the five highest flow volume years (High Flows) for total uncertainty generated from COP selected simulations. Cumulative simulation density is shaded for each day with black representing zero density and white representing area above the highest flow simulation. x-axis labels are simplified to only show the first day of each month. DOI: https://doi.org/10.1525/elementa.431.f5
be better defined. While informal likelihood functions are used in this study, the use of formal likelihood functions would not reject simulations, but rather assign them to lower likelihoods (Beven and Binley, 2014). Therefore, the inclusion of the distribution tails is present in most uncertainty studies, but they are simply rejected in favor of narrower uncertainty bounds (Ajami et al., 2007).

### 6.2. Interaction of uncertainties

The interaction of uncertainty sources was impacted by the performance metric used to select simulations. Differences in performance metrics have been a relevant topic in recent literature, particularly for use in multi-objective calibration (e.g., Asadzadeh and Tolson, 2013). Other performance metrics that do not ingest hydrometric uncertainty, such as the Nash-Sutcliffe Efficiency, were considered as behavioral simulation criteria, but produced little variation in reliability and sharpness values. Ingesting hydrometric uncertainty with the \( F_{\text{hyd}} \) metric generally produced sharper ensembles but focused on increasing reliability for high flows (Westerberg et al., 2020). The COP selected simulation ensembles were generally the sharpest and included fewer simulations than the top 10% (Appendix A). Since OLHS is used, no single performance metric is the focus of the optimization. Many calibration frameworks exist that can likely offer efficient sampling (e.g., Efstratiadis and Koutsoyiannis, 2010; Shafii and Tolson, 2015), however, studies such as Ajami et al. (2007) show that input uncertainty affects parameter distributions. Separate calibrations for each input dataset would explore different areas of the response surface and narrow the range of parameter uncertainty, thereby limiting the exploration of low likelihood events.

Studies such as Vaze et al. (2010) suggest models are less able to simulate conditions that are wetter or drier than those a model is calibrated to. The results presented in Figure 3 agree with the results of Vaze et al. (2010) as they show low performance for very dry or wet inputs, but highlight the variation of structural and parametric uncertainty when biases are introduced through the input data. Parameter uncertainty decreases with the ensemble minimum precipitation input and increases with the ensemble maximum. An interpretation of this is that when precipitation volume is sufficiently low, most water is lost to infiltration or evapotranspiration (ET), with remaining soil reservoir capacity or potential ET (PET) capacity available. Changing the capacities of PET and soil reservoirs would affect fewer events, which would lower sensitivity. Similarly, when precipitation volume is sufficiently high, PET and soil reservoirs are often exceeded, which suggests smaller parameter changes could be more impactful to a larger number of precipitation events (Vase et al., 2010; Chen et al., 2011; Merz et al., 2011; Coron et al., 2012; Brigode et al., 2013).

The propagation of structural uncertainty also displays distinct interaction with input uncertainty. Spatial aggregation, such as that of HEC-HMS, generally reduced the sensitivity to the ingested input uncertainty (Fischer et al., 2013; Pendergrass et al., 2017); this further suggests that model structure affects the sensitivity of climate change projections. In a data-sparse region, including low likelihood uncertainty distribution tails may also improve the ability to explore changes in less frequent events, such as floods or droughts (Mendoza et al., 2016). Under climate change, these distribution tails may be more sensitive to climate shifts (Vaze et al., 2010; Mendoza et al., 2016). Therefore, consideration of cumulative sources of uncertainty that extend into low likelihood uncertainty distribution tails can be considered a viable method for examining different percentile ranges of propagated uncertainty towards understanding low likelihood events in a data-sparse region.

### 7. Conclusions

Propagated uncertainty from input, parameter, and structural uncertainty sources for an ensemble of hydrologic models are generated and evaluated against a simple estimate of hydrometric uncertainty for several hydrometric flow data gauges. Results are highlighted as follows:

- Structural uncertainty without the inclusion of hydrometric uncertainty will likely appear as the highest quality flow ensemble;
- When compared with hydrometric uncertainty, neither parameter nor structural uncertainties alone are sufficient to represent the range of hydrometric uncertainty;
- Inclusion of all sources of uncertainty generates the widest uncertainty bounds, but the highest reliability;
- The relative magnitude of structural and parameter uncertainty is a function of input uncertainty, suggesting variations in the contribution to total uncertainty by source, with increased variability expected for climate change studies and low likelihood events; and
- The generation of wide uncertainty bounds that include low likelihood simulations can be used to explore low likelihood streamflow events in a data-sparse environment.

Consideration of all sources of uncertainty is expected to improve the quality of the simulated ensemble although the generated uncertainty bounds widen into low likelihood uncertainty distribution tails if simulations are not selected with ensemble optimization techniques, which offers benefits for data-sparse environments. Each source of uncertainty adds further value, but, without the consideration of hydrometric uncertainty, the added value may not be obvious. Basin scale effects the sources of uncertainty and how they propagate; additional research into the cumulative effects of uncertainty propagation at different basin scales should be considered. Additionally, further research into the evolution of structural and parameter uncertainty when forced by GCM climate projections should be considered, since the contribution of each source of uncertainty displayed interaction with input data and low likelihood uncertainty distribution tails are likely to be sensitive to climatic shifts.
Appendix A

Figure A1: Reliability and sharpness partitioned in 10th percentile binned increments using Fojb selected simulations. The logarithms of sharpness values are presented to increase comparability across results. Reliability and sharpness scores were calculated on daily timesteps for the full 1981–2010 period. DOI: https://doi.org/10.1525/elementa.431.a1

Figure A2: Reliability and sharpness partitioned in 10th percentile binned increments using COP selected simulations. The logarithms of Sharpness values are presented to increase comparability across results. Reliability and sharpness scores were calculated on daily timesteps for the full 1981–2010 period. DOI: https://doi.org/10.1525/elementa.431.a2
Data Accessibility Statement
Meteorological data used in this study are publically available datasets. All hydrologic outputs are publicly available. Figures following the format of Figure 3 and Figure 4 are also available in the OSF data repository. 10.17605/OSF.IO/7Z3J6.

Acknowledgements
Thanks to the University of Manitoba for supporting this research. Thanks to developers of the gridded datasets and hydrologic models used in this study. A special thanks to Tegan Holmes and Andrew Tefs for training and support for the hydrologic models used in this study. We would also like to thank the manuscript reviewers for their dedication to the improvement of this manuscript.

Funding information
This research was funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada, a national granting agency. Research was conducted as part of the BaySys collaborative research and Development NSERC grant (D. Barber), which was jointly supported by Manitoba Hydro (cash contribution) and the following agencies: ArcticNet, Ouranos, Hydro Quebec, and various Canadian Universities (in-kind).

Competing interests
The authors have no competing interests to declare.

Author contributions
- Conception and design: SP, TAS, GA, RL, SJD
- Acquisition of data: SP, RL
- Analysis and interpretation of data: SP, TAS, GA
- Drafted the article: SP, TAS, GA
- Review and revision of the article: SP, TAS, GA, RL, SJD, KK
- Approved the submitted version for publication: SP, TAS, GA, RL, SJD

References
Abebe, NA, Ogden, FL and Pradhan, NR. 2010. Sensitivity and uncertainty analysis of the conceptual HBV rainfall–runoff model: implications for parameter estimation. Journal of Hydrology 389: 301–310. DOI: https://doi.org/10.1016/j.jhydrol.2010.06.007

Addor, N, Jaun, S, Fundel, F and Zappa, M. 2011. An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios. Hydrology and Earth System Sciences 15(7): 2327–2347. DOI: https://doi.org/10.5194/hess-15-2327-2011

Ajami, NK, Duan, Q and Sorooshian, S. 2007. An integrated hydrologic bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resources Research 43(1): Article W01403. DOI: https://doi.org/10.1029/2005WR004745

Asadzadeh, M and Tolson, B. 2013. Pareto archived dynamically dimensioned search with hypervolume-based selection for multi-objective optimization. Engineering Optimization 45(12): 1489–1509. DOI: https://doi.org/10.1080/0305215X.2012.748046

Berg, P, Donnelly, C and Gustafsson, D. 2018. Near-real-time adjusted reanalysis forcing data for hydrology. Hydrology and Earth System Sciences 22(2): 989–1000. DOI: https://doi.org/10.5194/hess-22-989-2018

Beven, K. 2006. A manifesto for the equifinality thesis. Journal of Hydrology 320(1–2): 18–36. DOI: https://doi.org/10.1016/j.jhydrol.2005.07.007

Beven, K. 2016. Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication. Hydrological Sciences Journal 61(9): 1652–1665. DOI: https://doi.org/10.1080/0305215X.2015.1031761

Beven, K and Binley, A. 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 6(3): 279–298. DOI: https://doi.org/10.1002/hyp.360060305

Beven, K and Binley, A. 2014. GLUE: 20 years on. Hydrological Processes 28(24): 5897–5918. DOI: https://doi.org/10.1002/hyp.10082

Bourgin, F, Andréassian, V, Perrin, C and Oudin, L. 2015. Transferring global uncertainty estimates from gauged to ungauged catchments. Hydrology and Earth System Sciences 19: 2535–2546. DOI: https://doi.org/10.5194/hess-19-2535-2015

Brigode, P, Oudin, L and Perrin, C. 2013. Hydrological model parameter instability: A source of additional uncertainty in estimating the hydrological impacts of climate change? Journal of Hydrology 476: 410–425. DOI: https://doi.org/10.1016/j.jhydrol.2012.11.012

Brown, JD and Heuvelink, GB. 2006. Assessing uncertainty propagation through physically based models of soil water flow and solute transport. Encyclopedia of Hydrological Sciences, 1181–1195. John Wiley and Sons. DOI: https://doi.org/10.1002/0470848944.hsa081

Chen, J, Brissette, FP, Poulin, A and Leconte, R. 2011. Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. Water Resources Research 47(12): Article W12509. DOI: https://doi.org/10.1029/2011WR010602

Choi, W, Kim, SJ, Rasmussen, PF and Moore, AR. 2009. Use of the North American regional reanalysis for hydrological modeling in Manitoba. Canadian Water Resources Journal 34(1): 17–36. DOI: https://doi.org/10.4296/cwrj3401017

Clark, MP, Kavetski, D and Fencia, F. 2011. Pursuing the method of multiple working hypotheses for hydrological modeling. Water Resources Research 47(9): W09301. DOI: https://doi.org/10.1029/2010WR009827

Clark, MP, Nijssen, B, Lundquist, JD, Kavetski, D, Rupp, DE, Woods, RA, Freer, JE, Gutmann, ED, Wood, AW, Brekke, LD, Arnold, JR, Gochis, DJ and Rasmussen, RM. 2015. A unified approach for process-based hydrologic modeling: 1. Modeling
concept. Water Resources Research 51(4):2498–2514. DOI: https://doi.org/10.1002/2015WR017198

Cloke, HL and Pappenberger, F. 2009. Ensemble flood forecasting: A review. Journal of Hydrology 375(3–4): 613–626. DOI: https://doi.org/10.1016/j.jhydrol.2009.06.005

Coron, L, Andreassian, V, Perrin, C, Lerat, J, Vaze, J, Bourqui, M and Hendrickx, F. 2012. Crash testing hydrological models in contrasted climate conditions: an experiment on 216 Australian catchments. Water Resources Research 48(5): Article W05552. DOI: https://doi.org/10.1029/2011WR011721

Coxon, G, Freer, J, Westerberg, IK, Wagener, T, Woods, R and Smith, PJ. 2015. A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations. Water Resources Research 51(6): 531–5546. DOI: https://doi.org/10.1002/2014WR016532

Dams, J, Nosent, J, Senbeta, TB, Willems, P and Batelaan, O. 2015. Multi-model approach to assess the impact of climate change on runoff. Journal of Hydrology 529: 1601–1616. DOI: https://doi.org/10.1016/j.jhydrol.2015.08.023

Dee, DP, Uppala, SM, Simmons, AJ, Berrisford, P, Poli, P, Kobayashi, S, Andrae, U, Balmaseda, MA, Balsamo, G, Bauer, P, Bechtold, P, Beljaars, ACM, Van de Berg, L, Bidlot, J, Bormann, N, Delsol, C, Dragnani, R, Fuentes, M, Geer, AJ, Haimberger, L, Healy, SB, Hersbach, H, Holm, EV, Isaksen, L, Källberg, P, Köhler, M, Matricardi, M, McNally, AP, Monge-Sanz, BM, Morcrette, JJ, Park, BK, Peubey, C, de Rosnay, P, Tavolato, C, Thépaut, JN and Vitart, F. 2011. The ERA-Interim Reanalysis: configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society 137(656): 553–597. DOI: https://doi.org/10.1002/qj.828

Demeritt, D, Cloke, H, Pappenberger, F, Thielen, J, Bartholmes, J and Ramos, MH. 2007. Ensemble predictions and perceptions of risk, uncertainty, and error in flood forecasting. Environmental Hazards 7(2): 115–127. DOI: https://doi.org/10.1016/j.envhaz.2007.05.001

Dingman, SL. 2015. Physical Hydrology (3rd Edition). Waveland Press.

Dwarakish, GS and Ganarski, BP. 2015. Impact of land use change on hydrological systems: A review of current modeling approaches. Cogent Geoscience 1(1): 1115691. DOI: https://doi.org/10.1080/23312041.2015.1115691

Eckhardt, K, Breuer, L and Frede, HG. 2003. Parameter uncertainty and the significance of simulated land use change effects. Journal of Hydrology 273(1–4): 164–176. DOI: https://doi.org/10.1016/S0022-1694(02)00395-5

Efstratiadis, A and Koutsoyiannis, D. 2010. One decade of multi-objective calibration approaches in hydrological modelling: A review. Hydrological Sciences Journal 55(1): 58–78. DOI: https://doi.org/10.1080/02626660903526292

Environment Canada. 1980. Manual of hydrometric data computation and publication procedures, Fifth Edition, Inland Waters Directorate, Water Resources Branch, Ottawa.

Environment and Climate Change Canada. 2019. HYDAT version 2019-2.18. Surface water and sediment data. Water Survey of Canada, Environment and Climate Change Canada.

Eum, HI, Dibike, Y, Prowse, T and Bonsal, B. 2014. Inter-comparison of high-resolution gridded climate data sets and their implication on hydrological model simulation over the Athabasca watershed, Canada. Hydrological Processes 28(14): 4250–4271. DOI: https://doi.org/10.1002/hyp.10236

Fischer, EM, Beyerle, U and Knutti, R. 2013. Robust spatially aggregated projections of climate extremes. Nature Climate Change 3(12): 1033–1038. DOI: https://doi.org/10.1038/nclimate2051

Gbambie, AS, Poulin, A, Boucher, MA and Arsenault, R. 2017. Added value of alternative information in interpolated precipitation datasets for hydrology. Journal of Hydroeteorology 18(1): 247–264. DOI: https://doi.org/10.1175/JHM-D-16-0032.1

Gneiting, T, Balabdaoui, F and Raftery, AE. 2007. Probabilistic forecasts, calibration and sharpness. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69(2): 243–268. DOI: https://doi.org/10.1111/j.1467-9868.2007.00587.x

Gupta, HV, Kling, H, Yilmaz, KK and Martinez, GF. 2009. Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modelling. Journal of Hydrology 377(1–2): 80–91. DOI: https://doi.org/10.1016/j.jhydrol.2009.08.003

Hamilton, S. 2008. Sources of uncertainty in Canadian low flow hydrometric data. Canadian Water Resources Journal 33(2): 125–136. DOI: https://doi.org/10.4296/cwj3302125

Hamilton, AS and Moore, RD. 2012. Quantifying uncertainty in streamflow records. Canadian Water Resources Journal 37(1): 3–21. DOI: https://doi.org/10.4296/cwj3701865

Han, S and Coulibaly, P. 2019. Probabilistic flood forecasting using hydrologic uncertainty processor with ensemble weather forecasts. Journal of Hydroeteorology 20(7): 1379–1398. DOI: https://doi.org/10.1175/JHM-D-18-0251.1

Holmes, T. 2016. Assessing the value of stable water isotopes in hydrologic modeling: A dual-isotope approach. MSc thesis, University of Manitoba.

Huard, D and Mailhot, A. 2006. A Bayesian perspective on input uncertainty in model calibration: Application to hydrological model “abc”. Water Resources Research 42(7): W07416. DOI: https://doi.org/10.1029/2005WR004661

Hubbard, KG, Mahmood, R and Carlson, C. 2003. Estimating daily dew point temperature for the northern Great Plains using maximum and minimum temperature. Agronomy Journal 95(2): 323–328. DOI: https://doi.org/10.2134/agronj2003.3230
Hutchinson, MF, McKenney, DW, Lawrence, K, Pedlar, JH, Hopkinson, RF, Milewska, E and Papadopoul, P. 2009. Development and testing of Canada-wide interpolated spatial models of daily minimum–maximum temperature and precipitation for 1961–2003. Journal of Applied Meteorology and Climatology 48(4): 725–741. DOI: https://doi.org/10.1175/2008JAMC1979.1

Karlsso, IB, Sonnenborg, TO, Refsgaard, JC, Trolle, D, Børgezen, CD, Olesen, JE and Jensen, KH. 2016. Combined effects of climate models, hydrological model structures and land use scenarios on hydrological impacts of climate change. Journal of Hydrology 535: 301–317. DOI: https://doi.org/10.1016/j.jhydrol.2016.01.069

Kavetski, D, Franks, SW and Kuczera, G. 2003. Confronting input uncertainty in environmental modelling. Calibration of Watershed Models 6: 49–68. DOI: https://doi.org/10.1029/WO006p0049

Kavetski, D, Kuczera, G and Franks, SW. 2006. Bayesian analysis of input uncertainty in hydrological modeling: I. Theory. Water Resources Research 42(3): W03407. DOI: https://doi.org/10.1029/2005WR004368

Kiang, JE, Gazoorian, C, McMillan, H, Coxon, G, Le Coz, J, Westerberg, IK, Belleville, A, Sevrez, D, Sikorska, AE, Petersen-Øverleir, A, Reitan, T, Freer, J, Renard, B, Mansanarez, V and Mason, R. 2018. A comparison of methods for streamflow uncertainty estimation. Water Resources Research 54(10): 7149–7176. DOI: https://doi.org/10.1002/2018WR022708

Kouwen, N. 2018. WATFLOOD users manual. Water Resources Group, University of Waterloo.

Li, L, Xia, J, Xu, CY and Singh, VP. 2010. Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. Journal of Hydrology 390(3–4): 210–221. DOI: https://doi.org/10.1016/j.jhydrol.2010.06.044

Li, L and Xu, CY. 2014. The comparison of sensitivity analysis of hydrological uncertainty estimates by GLUE and Bayesian method under the impact of precipitation errors. Stochastic Environmental Research and Risk Assessment 28(3): 491–504. DOI: https://doi.org/10.1007/s00477-013-0767-1

Lilhare, R, Déry, SJ, Pokorny, S, Stadnyk, TA and Koenig, KA. 2019. Intercomparison of multiple hydroclimatic datasets across the Lower Nelson River Basin, Manitoba, Canada. Atmosphere-Ocean 57: 262–278. DOI: https://doi.org/10.1080/07055900.2019.1638226

Matott, LS, Babendreier, JE and Purucker, ST. 2009. Evaluating uncertainty in integrated environmental models: a review of concepts and tools. Water Resources Research 45(6): Article W06421. DOI: https://doi.org/10.1029/2008WR007301

McMillan, H, Freer, J, Pappenberger, F, Krueger, T and Clark, M. 2010. Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. Hydrological Processes 24(10): 1270–1284. DOI: https://doi.org/10.1002/hyp.7587

McMillan, H, Jackson, B, Clark, M, Kavetski, D and Woods, R. 2011. Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. Journal of Hydrology 400(1–2): 83–94. DOI: https://doi.org/10.1016/j.jhydrol.2010.01.026

McMillan, H, Krueger, T and Freer, J. 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrological Processes 26(26): 4078–4111. DOI: https://doi.org/10.1002/hyp.9384

McMillan, HK, Westerberg, IK and Krueger, T. 2018. Hydrological data uncertainty and its implications. Wiley Interdisciplinary Reviews: Water 5(6), e1319. DOI: https://doi.org/10.1002/wat2.1319

Mei, Y, Nikoloopoulos, EI, Anagnostou, EN and Borga, M. 2016. Evaluating satellite precipitation error propagation in runoff simulations of mountainous basins. Journal of Hydrometeorology 17(5): 1407–1423. DOI: https://doi.org/10.1175/JHM-D-15-0081.1

Mendoza, PA, Clark, MP, Mizukami, N, Gutmann, ED, Arnold, JR, Brekke, LD and Rajagopalan, B. 2016. How do hydrologic modeling decisions affect the portrayal of climate change impacts? Hydrological Processes 30(7): 1071–1095. DOI: https://doi.org/10.1002/hyp.10684

Merz, R, Parajka, J and Blöschl, G. 2011. Time stability of catchment model parameters: Implications for climate impact analyses. Water Resources Research 47(2): Article W02531. DOI: https://doi.org/10.1029/2010WR009505

Mesinger, F, DiMego, G, Kalnay, E, Mitchell, K, Shafran, PC, Ebisuzaki, W, Jović, D, Woollen, J, Rogers, E, Berbery, EH, Ek, MB, Fan, Y, Grumbine, R, Higgins, W, Li, H, Lin, Y, Manikin, G, Parrish, D and Shi, W. 2006. North American regional reanalysis. Bulletin of the American Meteorological Society 87(3): 343–360. DOI: https://doi.org/10.1175/BAMS-87-3-343

Montanari, A. 2005. Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations. Water Resources Research 41(8): W08406. DOI: https://doi.org/10.1029/2004WR003826

Montanari, A and Di Baldassarre, G. 2013. Data errors and hydrological modelling: The role of model structure to propagate observation uncertainty. Advances in Water Resources 51: 498–504. DOI: https://doi.org/10.1016/j.advwatres.2012.09.007

Muhammad, A, Evenson, GR, Stadnyk, TA, Boluwade, A, Jha, SK and Coulilby, P. 2018a. Assessing the importance of potholes in the Canadian Prairie Region under future climate change scenarios. Water 10(11): 1657. DOI: https://doi.org/10.3390/w10111657

Muhammad, A, Evenson, GR, Stadnyk, TA, Boluwade, A, Jha, SK and Coulilby, P. 2019. Impact of model structure on the accuracy of hydrological
modeling of a Canadian Prairie watershed. Journal of Hydrology: Regional Studies 21: 40–56. DOI: https://doi.org/10.1016/j.jheirh.2018.11.005

Muhammad, A, Stadnyk, T, Unduche, F and Coulibaly, P. 2018b. Multi-model approaches for improving seasonal ensemble streamflow prediction scheme with various statistical post-processing techniques in the Canadian Prairie region. Water 10(11): 1604. DOI: https://doi.org/10.3390/w10111604

Nikolopoulos, El, Anagnostou, EN, Hossain, F, Gebremichael, M and Borga, M. 2010. Understanding the scale relationships of uncertainty propagation of satellite rainfall through a distributed hydrologic model. Journal of Hydro meteorology 11(2): 520–532. DOI: https://doi.org/10.1175/2009JHM1169.1

Pappenberger, F, Stephens, E, Thielen, J, Salamon, P, Demeritt, D, Jan van Andel, S, Wetterhall, F and Alfieri, L. 2013. Visualizing probabilistic flood forecast information: expert preferences and perceptions of best practice in uncertainty communication. Hydro logical Processes 27(1): 132–146. DOI: https://doi.org/10.1002/hyp.9253

Pechlivanidis, IG, Jackson, BM, McIntyre, NR and Wheeler, HS. 2011. Catchment scale hydrological modelling: a review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. Global NEST Journal 13(3): 193–214. DOI: https://doi.org/10.3905/gnj.000778

Pendergrass, AG, Knutti, R, Lehner, F, Deser, C and Sanderson, BM. 2017. Precipitation variability increases in a warmer climate. Scientific Reports 7(1): 17966. DOI: https://doi.org/10.1038/s41598-017-17966-y

Pokorny, S. 2019. Assessing the relative contributions of input, structural, parameter, and output uncertainties to total uncertainty in hydrologic modeling. MSc thesis, University of Manitoba.

Priestley, CHB and Taylor, RJ. 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Monthly Weather Review 100(2): 81–92. DOI: https://doi.org/10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2

Rapaic, M, Brown, R, Markovic, M and Chaumont, D. 2015. An evaluation of temperature and precipitation surface-based and reanalysis datasets for the Canadian Arctic, 1950–2010. Atmosphere-Ocean 53(3): 283–303. DOI: https://doi.org/10.1080/07055900.2015.1045825

Renard, B, Kavetski, D, Kuczera, G, Thyer, M and Franks, SW. 2010. Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. Water Resources Research 46(5): W05521. DOI: https://doi.org/10.1029/2009WR008328

Rokaya, P, Budhathoki, S and Lindenschmidt, KE. 2018. Trends in the timing and magnitude of ice-jam floods in Canada. Scientific Reports 8(1): 5834. DOI: https://doi.org/10.1038/s41598-018-24057-z

Sagan, KAB. 2017. Sensitivity of probable maximum flood estimates in the Lower Nelson River Basin. MSc thesis, University of Manitoba.

Shafii, M, Tolson, B and Matott, LS. 2015. Addressing subjective decision-making inherent in GLUE-based multi-criteria rainfall–runoff model calibration. Journal of Hydrology 523: 693–705. DOI: https://doi.org/10.1016/j.jhydrol.2015.01.051

Shilkomanov, AI, Yakovleva, TI, Lammers, RB, Karasev, IP, Vörösmarty, CJ and Linder, E. 2006. Cold region river discharge uncertainty—Estimates from large Russian rivers. Journal of Hydrology 326(1–4): 231–256. DOI: https://doi.org/10.1016/j.jhydrol.2005.10.037

SMHI. 2018. HYPE model documentation. Retrieved from: http://www.smhi.net/hype/wiki/doku.php

Smith, A. 2015. Utilizing lumped coupled tracer-aided modelling to identify temporal trends in basin-scale evapotranspiration partitioning. MSc thesis, University of Manitoba.

Stadnyk, TA, MacDonald, M, Tefs, A, Awoye, H, Dery, SJ, Gustafsson, D, Isberg, K and Arheimer, B. Accepted. Freshwater discharge into Hudson Bay Simulated by HYPE. Elementa: Science of the Anthropocene. EMID:401da07a0557532f.

Stedinger, JR, Vogel, RM, Lee, SU and Batchelder, R. 2008. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. Water Resources Research 44(12): Article W00B06. DOI: https://doi.org/10.1029/2008WR006822

Tang, B. 1993. Orthogonal array-based Latin hypercubes. Journal of the American Statistical Association 88(424): 1392–1397. DOI: https://doi.org/10.1080/01621459.1993.10476423

Tasdighi, A, Arabi, M and Harmel, D. 2018. A probabilistic appraisal of rainfall-runoff modeling approaches within SWAT in mixed land use watersheds. Journal of Hydrology 564: 476–489. DOI: https://doi.org/10.1016/j.jhydrol.2018.07.035

USACE. 2016. Hydrologic modeling system HEC-HMS. User’s manual.

Uusitalo, I, Lehikoinen, A, Helle, I and Myrberg, K. 2015. An overview of methods to evaluate uncertainty of deterministic models in decision support. Environmental Modelling and Software 63: 24–31. DOI: https://doi.org/10.1016/j.envsoft.2014.09.017

Vaze, J, Post, DA, Chiew, FHS, Perraud, JM, Viney, NR and Teng, J. 2010. Climate non-stationarity—validity of calibrated rainfall–runoff models for use in climate change studies. Journal of Hydrology 394(3–4): 447–457. DOI: https://doi.org/10.1016/j.jhydrol.2010.09.018

Vrugt, JA, Diks, CG, Gupta, HV, Boutsen, W and Verstraten, JM. 2005. Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. Water Resources Research 41(1): W01017. DOI: https://doi.org/10.1029/2004WR003059

Vrugt, JA, Ter Braak, CJ, Clark, MP, Hyman, JM and Robinson, BA. 2008. Treatment of input uncertainty
in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research 44(12): W00809. DOI: https://doi.org/10.1029/2007WR006720

Wagener, T, McIntyre, N, Lees, MJ, Wheater, HS and Gupta, HV. 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. Hydrological Processes 17(2): 455–476. DOI: https://doi.org/10.1002/hyp.1135

Weedon, GP, Balsamo, G, Bellouin, N, Gomes, S, Best, MJ and Viterbo, P. 2014. The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. Water Resources Research 50(9): 7505–7514. DOI: https://doi.org/10.1002/2014WR015638

Westerberg, IK, Guerrero, JL, Younger, PM, Beven, KJ, Seibert, J, Halldin, S, Freer, JE and Xu, CY. 2011. Calibration of hydrological models using flow-duration curves. Hydrology and Earth System Sciences 15(7): 2205–2227. DOI: https://doi.org/10.5194/hess-15-2205-2011

Westerberg, IK, Sikorska-Senoner, AE, Viviroli, D, Vis, M and Seibert, J. 2020. Hydrological model calibration with uncertain discharge data. Hydrological Sciences Journal: 1–16. DOI: https://doi.org/10.1080/02626667.2020.1735638

Westerberg, IK, Wagener, T, Coxon, G, McMillan, HK, Castellarin, A, Montanari, A and Freer, J. 2016. Uncertainty in hydrological signatures for gauged and ungauged catchments. Water Resources Research 52(3): 1847–1865. DOI: https://doi.org/10.1002/2015WR017635

Westmacott, JR and Burn, DH. 1997. Climate change effects on the hydrologic regime within the Churchill-Nelson River Basin. Journal of Hydrology 202(1–4): 263–279. DOI: https://doi.org/10.1016/S0022-1694(97)00073-5

Whitfield, PH and Pomeroy, JW. 2017. Assessing the quality of the streamflow record for a long-term reference hydrometric station: Bow River at Banff. Canadian Water Resources Journal 42(4): 391–415. DOI: https://doi.org/10.1080/07011784.2017.1399086

Wi, S, Yang, YCE, Steinschneider, S, Khalil, A and Brown, CM. 2015. Calibration approaches for distributed hydrologic models in poorly gaged basins: implication for streamflow projections under climate change. Hydrology and Earth System Sciences 19(2): 857–876. DOI: https://doi.org/10.5194/hess-19-857-2015

Wong, JS, Razavi, S, Bonsal, BR, Wheater, HS and Asong, ZE. 2017. Inter-comparison of daily precipitation products for large-scale hydro-climatic applications over Canada. Hydrology and Earth System Sciences 21(4): 2163–2185. DOI: https://doi.org/10.5194/hess-21-2163-2017

Yadav, M, Wagener, T and Gupta, H. 2007. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. Advances in Water Resources 30(8): 1756–1774. DOI: https://doi.org/10.1016/j.advwatres.2007.01.005

Zhou, R, Li, Y, Lu, D, Liu, H and Zhou, H. 2016. An optimization based sampling approach for multiple metrics uncertainty analysis using generalized likelihood uncertainty estimation. Journal of Hydrology 540: 274–286. DOI: https://doi.org/10.1016/j.jhydrol.2016.06.030

How to cite this article: Pokorny, S, Stadnyk, TA, Ali, G, Lihare, R, Déry, SJ and Koenig, K. 2021. Cumulative Effects of Uncertainty on Simulated Streamflow in a Hydrologic Modeling Environment. Elem Sci Anth, 9: 1. DOI: https://doi.org/10.1525/elementa.431

Domain Editor-in-Chief: Steven Allison, University of California Irvine, US

Associate Editor: Julian D. Olden, School of Aquatic & Fishery Sciences, University of Washington, US

Knowledge Domain: Ecology and Earth Systems

Part of an Elementa Special Feature: BaySys

Submitted: 21 February 2020 Accepted: 22 June 2020 Published: 13 January 2021

Copyright: © 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See http://creativecommons.org/licenses/by/4.0/.

Elem Sci Anth is a peer-reviewed open access journal published by University of California Press.