


## Article

# Identification of Hydrological Models for Enhanced Ensemble Reservoir Inflow Forecasting in a Large Complex Prairie Watershed

Frezer Seid Awol <sup>1,\*</sup> , Paulin Coulibaly <sup>2</sup>, Ioannis Tsanis <sup>1,3</sup> and Fisaha Unduche <sup>4</sup>

<sup>1</sup> Department of Civil Engineering, McMaster University, 1280 Main Street West, Hamilton, ON L8S4L7, Canada; tsanis@hydromech.gr

<sup>2</sup> Jointly in School of Geography and Earth Sciences and Department of Civil Engineering, McMaster University, 1280 Main Street West, Hamilton, ON L8S4K1, Canada; couliba@mcmaster.ca

<sup>3</sup> School of Environmental Engineering, Technical University of Crete, 73100 Chania, Greece

<sup>4</sup> Hydrologic Forecasting & Coordination, Manitoba Infrastructure, 280 Broadway, Winnipeg, MB R3C 0R8, Canada; fisaha.unduche@gov.mb.ca

\* Correspondence: awolf@mcmaster.ca

Received: 11 September 2019; Accepted: 18 October 2019; Published: 23 October 2019



**Abstract:** Accurate and reliable flow forecasting in complex Canadian prairie watersheds has been one of the major challenges faced by hydrologists. In an attempt to improve the accuracy and reliability of a reservoir inflow forecast, this study investigates structurally different hydrological models along with ensemble precipitation forecasts to identify the most skillful and reliable model. The key goal is to assess whether short- and medium-range ensemble flood forecasting in large complex basins can be accurately achieved by simple conceptual lumped models (e.g., SACSMA with SNOW17 and MACHBV with SNOW17) or it requires a medium level distributed model (e.g., WATFLOOD) or an advanced macroscale land-surface based model (VIC coupled with routing module (RVIC)). Eleven (11)-member precipitation forecasts from second-generation Global Ensemble Forecast System reforecast (GEFSv2) were used as inputs. Each of the ensemble members was bias-corrected by Empirical Quantile Mapping method using the Canadian Precipitation Analysis (CaPA) as a training/verification dataset. Forecast evaluation is performed for 1-day up to 8-days forecast lead times in a 6-month hindcast period. Results indicate that bias-correcting precipitation forecasts using verifying datasets (such as CaPA) for a training period of at least two years before the forecast time, produces skillful ensemble hydrological forecasts. A comparison of models in forecast mode shows that the two lumped models (SACSMA and MACHBV) can provide better overall forecast performance than the benchmark WATFLOOD and the macroscale Variable Infiltration Capacity (VIC) model. However, for shorter lead-times, particularly up to day 3, the benchmark distributed model provides competitive reliability, as compared to the lumped models. In general, the SACSMA model provided better forecast quality, reliability and differentiation skill than other considered models at all lead times.

**Keywords:** hydrological models; ensemble hydrological forecasting; bias-correction; SACSMA; complex watersheds; reservoir inflow

## 1. Introduction

Prairie watersheds are characterized by several small depressions, potholes, and wetlands, and poorly connected drainage systems that may or may not contribute to the main river system [1]. They are often featured by their long winter periods, high spring snowmelt contribution to annual runoff, deep-frozen soils and rapid infiltration, intense rainfall in spring and early summer, lower

soil moisture, and evaporation from summer to fall [1]. Relevant methodologies were proposed to assess several aspects of the hydrological cycle such as snowpack, spring melt, soil moisture, rainfall frequency, and evaporation, in the Canadian Prairie regions [2–5]. The effect of climate, land use, and ecosystem change on the hydrological processes of cold and wetland regions were also studied [6–8]. Even though some efforts were made to formulate the realistic representation of wetland processes in hydrological models [9–13], challenges of hydrological forecasting and flood predictions in such complex watersheds remain at large.

Several important works have already been performed for enhancing flood prediction in several watersheds: for example, using single or multiple hydrological models [14–22], or feeding ensemble numerical weather products to models [23–29]. Velázquez et al. [18], for example, analyzed 16 lumped hydrological models with 50-member ensemble weather inputs. They detected that the multi-model approach of a grand member ensemble provided more forecast skill and reliability than either a single model with meteorological ensembles or multiple models with the deterministic forecast at all lead times. Pietroniro et al. [27], assessed the benefit of using Environment Canada’s MESH (Modélisation Environnementale Communautaire-MEC Surface and Hydrology) model in the Great Lakes catchment with inputs from 16-member ensemble forecast variables supplied by Meteorological Service of Canada (MSC). Fan et al. [30], suggested the use of local or regional ensemble forecasts instead of low-resolution global ensemble inputs, and data assimilation methods. In their work, they applied MGB-IHB distributed model with bias-corrected second-generation Global Ensemble Forecast System (GEFS v2) reforecast inputs and suggested that the improvements made could address the lack of spread in reservoir inflow forecasts especially in early lead times. Using the same hydrological model, Fan et al. [31], evaluated the importance of three sets of ensemble QPFs from the TIGGE (THORPEX Interactive Grand Global Ensemble) database in larger basins that have major reservoirs and hydroelectric plants. Their verification methods confirmed that the performance of hydrological forecasts depends on the quality of each ensemble precipitation products, but they also highlighted the improved reliability and robustness of ensemble river flows obtained from the combined super ensemble inputs. Abaza et al. [32], compared currently available Canadian meteorological forecasts and concluded that streamflow forecasting fed by Regional ensemble prediction systems (EPS) provided higher reliability than the Global EPS followed by their deterministic counterparts, as also supported by Fan et al. [30]. The use of multiple models with the global, regional, and local ensemble and deterministic inputs has also been implemented in several operational flood forecasting centers across the globe [33–39].

The main challenge in getting accurate and reliable short- and medium-range flood forecasts in large complex watersheds arises from the type of hydrological models, and the quality of weather forecast inputs applied. The choice of the models to be implemented for flood and streamflow forecasting depends on the intended purpose, the type of forecast inputs, and the complexity and scale of the study area [40]. Given the complexity of a prairie watershed in defining wetland and non-wetland physical processes and its representation by model structures for a specific application of real-time flood forecasting, it is essential to identify the candidate hydrological model(s) from multiple diverse potential models. Once the hydrological model or group of models are identified, the skill and reliability of hydrological forecasts can be enhanced by feeding qualitative ensemble weather forecasts into the models.

The limitations of previous works and the scientific challenges are that:

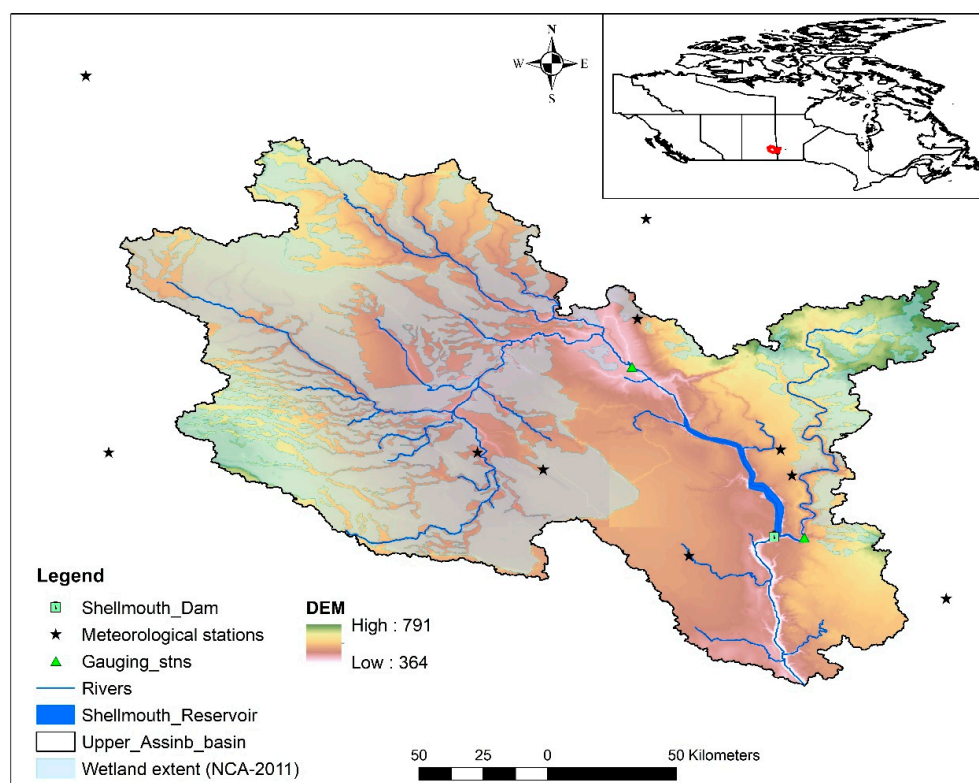
1. Only a few studies were conducted on large and complex Prairie watersheds,
2. Only a lumped model or a distributed model was used independently, for hydrological forecasting study. Alternatively, in some cases, the multi-models were only a collection of lumped conceptual models,
3. Identification of best hydrological model was usually based on historical meteorological or in some cases, deterministic weather forecast inputs. Evaluation and comparison of models based on raw and bias-corrected ensemble precipitation forecasts were not studied. As such, the objective

of this research is designed to address these limitations and identify the best hydrological model from diverse multi-models for short- and medium-range flood forecasting in a Canadian Prairie watershed. In this study, four structurally varied hydrological models were set up in order to simulate and forecast inflows to the Shelmouth Reservoir, which is located in Upper Assiniboine River Basin. A mixture of two lumped, one distributed and one macroscale land surface models were used in this research. In forecast mode, bias-corrected precipitation from second-generation Global Ensemble Forecast System (GEFS v2) reforecasts was fed into the four models in order to evaluate the reliability, skill, and overall forecast performance of the ensemble reservoir inflows.

## 2. Materials

### 2.1. Study Area

The Canadian Prairies are mainly located in Saskatchewan, Manitoba, and Alberta Provinces. The research is conducted in one of the main Canadian Prairie watersheds, the Upper Assiniboine River Basin upstream of the Shelmouth Reservoir, also called Lake of the Prairie (Figure 1). The catchment area contributing to the reservoir inflow is approximately 18,000 km<sup>2</sup>. While much of the basin is located in Saskatchewan, the Shelmouth Reservoir itself is located in the Province of Manitoba. Inflow into the reservoir is generated from three major upstream tributaries: the Whitesand River, the Shell Rivers and the main stream of the Assiniboine River. This Prairie watershed, which is known to have a complex hydrology is characterized by abundant potholes and wetlands, poorly interconnected streams and non-contributing areas, long and cold winter periods, deep-frozen soils and rapid infiltration, high spring snowmelt contribution to annual runoff, intense rainfall in spring and early summer, lower soil moisture and evaporation from summer to fall [1,39]. The basin's topography ranges from 250 m a.s.l. at its lowest point to 820 m a.s.l at its highest point, and its annual precipitation is approximately 460 mm [41]. The land cover of the basin is mostly dominated by cropland, which contributes about 55–58% of the land cover [41].



**Figure 1.** Study area of the Upper Assiniboine River Basin.

## 2.2. Data

### 2.2.1. Ensemble Weather Forecasts

An 11-member ensemble data from second-generation Global Ensemble Forecast System (GEFS v2) reforecast [42] supplied by National Centers for Environmental Prediction (NCEP), hereafter called “GEFSv2” was used as an input to models. The GEFSv2 issues forecast once a day in 3 hourly time step up to 8 days lead time with 50 km spatial resolution and the next eight days with a lower spatial resolution. For this research, daily total precipitation forecasts from January 2014 to December 2017 were used for the input datasets; the first two years used for bias correcting the last two years. Only precipitation forecasts were used as forcing data, while other variables were taken from observation because the accuracy of flood prediction is highly impacted by precipitation forecasts than any other variables [29].

### 2.2.2. Observed Data

Average daily temperature and precipitation data were obtained from Environment Canada for the eleven weather gauging stations that are distributed across the catchment (Figure 1). These data were used as inputs to the hydrological models. The output from the hydrological models, which is regarded as the simulated reservoir inflow was compared with calculated (observed) reservoir inflow in the calibration process. Detail information on the reservoir inflow estimation is provided in Section 2.2.3, whereas calibration and validation will be discussed in Section 3.2. In addition to the gauge data, precipitation data were also collected from The Canadian Precipitation Analysis (CaPA). The CaPA is developed by statistical interpolation of a background field from short-range precipitation forecasts, and observation from radar and ground-based rainfall measurements [43]. The spatial and temporal resolution of CaPA is 15 km and 6 h respectively. For this study, CaPA precipitation data is used for bias correcting global ensemble forecasts, which will be further discussed in Section 3.3.

### 2.2.3. Reservoir Inflow

The study area is the watershed upstream of the Shelmouth Reservoir. There is no flow gauge (actual streamflow measurement) at the outlet of the watershed. At the mouth of the reservoir or the Dam section, the outflow is regulated by structural mechanisms such as releasing water through the conduits (using gates) and spillways. These releases are controlled and measured daily. Therefore, the outflow from the reservoir is a regulated outflow measured at the conduits and spillway, and due to this reason, it cannot be directly used for calibration. Instead, the reservoir inflow is implicitly considered as the outflow from the entire watershed and is used for calibrating the hydrological models. The inflow is, in this case, a collection of water from major and minor tributaries that goes into the reservoir. The estimated inflow is considered as an “unregulated” discharge observation measuring collectively the river flows coming from the tributaries.

The inflow into Shemouth Reservoir is calculated based on a simple water balance equation. Given records of daily reservoir levels, the elevation-area-storage curve of the reservoir, and the summation of outflows measured at the spillway and conduit, the water balance can be formulated by the Equation (1). Here, losses (such as evaporation and infiltration) within a day are assumed to be negligible, and lateral inflows are included in ‘Inflow’ variable.

$$Inflow - Outflow = \frac{dS}{dt} \quad (1)$$

where  $dS/dt$  is the change in storage in a one-day time difference. The change in storage is obtained from the elevation-storage curve by looking at the daily average reservoir levels between the first and the second day. The reservoir inflow is calculated daily for practical application at Manitoba Hydrological Forecasting Center.

As described above, the reservoir inflow is regarded as a streamflow measurement of all the tributary rivers and streams supplying water to the reservoir. Since the inflow is not an actual flow measurement of the supplying rivers, it is prone to some degree of errors. However, the calculated reservoir inflow is believed to be the best possible method of measuring the “unregulated” watershed outflow. Also, there is uncertainty arising from the calculation method. We used a simple water balance equation to calculate the daily inflow, by only accounting for the daily change in storage and the daily measured regulated reservoir outflow. The daily losses (e.g., evaporation and infiltration) are assumed to negligible. Such an assumption might contain some uncertainties. However, the uncertainty for daily water balance is not believed to be considerable, for example, comparing with monthly water balance where such losses cannot be ignored.

### 3. Methods

Figure 2 shows the methodology adopted in this research. Ensemble weather forecasts products from GEFsV2 were collected. Each ensemble member of precipitation forecasts was bias-corrected by Empirical Quantile Mapping method using CaPA as a verifying data (Section 3.3). Four structurally various hydrological models were applied in the watershed including the benchmark model. Using raw and bias-corrected GEFsV2 ensemble inputs, the models’ forecasting performances were evaluated and compared in hindcast period.

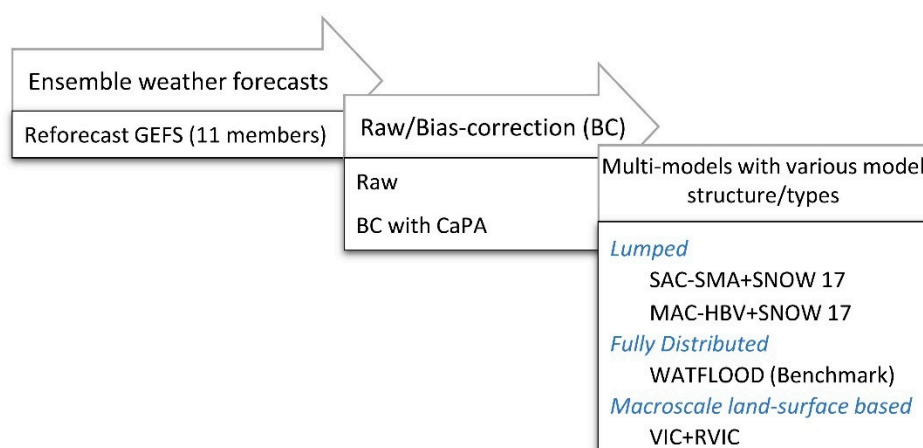


Figure 2. Illustration of the methodology adopted.

#### 3.1. Hydrological Models

Four different hydrological models with diverse model structures were applied in the study area to meet the research objectives described in Section 1. The models applied herein are a combination of two lumped models, one distributed and one macroscale land-surface based hydrological model. The lumped models are the Sacramento Soil Moisture Accounting (SAC-SMA) model coupled with SNOW17 [44] routine and the McMaster University-Hydrologiska Byråns Vattenbalansavdelning (MAC-HBV) [45] model coupled with SNOW17 routine. The third model applied is the macroscale land-surface based Variable Infiltration Capacity (VIC) model [46] coupled with a routing module. VIC has been applied in nearby, similar river basins for climate change and other hydrological studies [6,47–49]. The above three models were calibrated and validated in this research.

As a benchmark, we used the distributed WATFLOOD model [50]. For this study, a calibrated and operational WATFLOOD model was obtained from the Manitoba Infrastructure, Hydrological Forecasting Centre. The model has been used by the center to provide operational flood forecasting using real-time weather forecast data to issue short- and medium-range river forecasts in the Upper Assiniboine River basin and other nearby watersheds [39]. WATFLOOD is a Canadian Hydrological model specifically developed for flood forecasting and watershed simulation. The model is used



as a primary routing module for the Canadian national hydrological modeling system (MESH) [51]. Newman et al. [52], argues that a calibrated hydrological model which has a familiar practical application in local river forecasting systems, has a better functional capability than reference statistical systems to test models, and employs significant water budget interactions, is a suitable choice for use as a benchmark model.

For SACSMA and MACHBV models, mean areal daily temperature and precipitation time series data were created by using the Thiessen polygon method from eleven meteorological gauging stations that are distributed across the catchment. The average catchment elevation and latitude of the centroid values were used as an input to the SNOW17 model in addition to precipitation and temperature data. These inputs were used to calibrate and validate the two lumped models.

For the VIC (version 4.2.d) model, daily gridded interpolated precipitation data was generated from 11 gauging stations, using a bilinear interpolation technique. Daily gridded minimum and maximum temperature data were provided by the Natural Resources Canada, which applied a “thin-plate smoothing splines” (ANUSPLIN) method on observations from several ground-based stations in Canada to generate long-term daily gridded data [53,54]. ANUSPLIN has been used as forcing data for the VIC model in several studies [6,47–49]. Daily average wind speed data performed from the Global Environmental Multiscale (GEM) model [55]. The grid resolution of VIC model was about 1/8 degree. Land cover data is obtained from Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type (MCD12Q1) Version 6 data product [56]. Soil data were imported from FAO’s Harmonized World Soil Database V 1.2 [57]. The runoff from the land surface VIC grid cells was routed to and along the river networks using the routing module (RVIC) [58] based on Lohmann et al. [59].

For the WATFLOOD model, gridded interpolated daily precipitation and temperature data were used to set up and calibrate the model. The model was set up at a grid resolution of approximately 5 km for the Upper Assiniboine River Basin.

### 3.2. Calibration and Validation

The catchment outflow simulated by the hydrological models is considered as the reservoir inflow because the reservoir, located at the very downstream location, collects all water from tributary rivers and lateral inflows. During the calibration process, the comparison was made between the simulated and the observed daily reservoir inflow time series.

Dynamically Dimensioned Search (DDS) algorithm [60] was used to optimize the calibration of SACSMA/SNOW17, MACHBV/SNOW17, and VIC/RVIC models. Calibration and validation of the models were performed with daily timesteps from January 2005 to December 2015 with 1-year spin-up periods.

DDS has been previously compared with other optimization methods, such as shuffled complex evolution (SCE) by Tolson and Shoemaker [60]. In their study, the dimensionality and efficiency of DDS, for example, was tested, and the authors concluded that DDS provided better results than SCE both with low- and high-dimensional problems, and is more efficient. DDS has been used to calibrate several hydrological models from simple lumped to medium level distributed models (e.g., SWAT [61,62], MESH [63], CRHM-AHM [64]) to very complicated land-surface based models (e.g., WRF-Hydro [65,66]).

For lumped models, 10 parameters of SNOW17, 15 parameters of SACSMA, and 12 parameters of MACHBV were calibrated. The optimizing parameters are presented in Appendix A.1, Appendix A.2, and Appendix A.3 for SNOW17, SACSMA, and MACHBV, respectively. For VIC model, the total number of parameters to be optimized and calibrated is increased from the default 13 to 53 including the wetland and routing parameters. The optimizing parameters for the VIC model are presented in Appendix A.4. Similar to the lumped models, the simulated reservoir inflow from the VIC/RVIC model was compared with daily observed flow in the calibration process.

For all models, a single objective function obtained by a weighted average of two performance metrics was used in the DDS optimization. The performance metrics that were given equal weight are Kling–Gupta efficiency (KGE) [67], and Peak Flow Criteria (PFC) [68] as defined below.

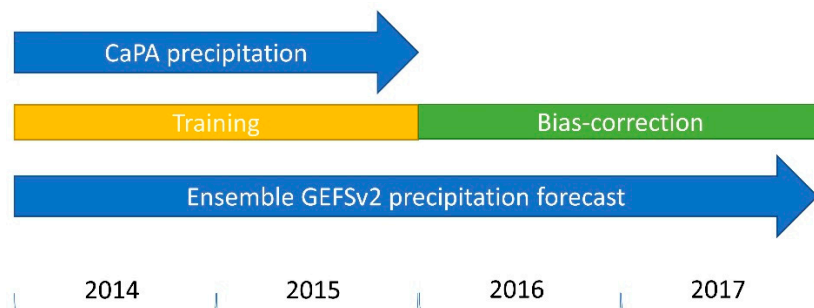
$$KGE = 1 - \sqrt{(r-1)^2 + (a-1)^2 + (b-1)^2} \quad (2)$$

$$PFC = \frac{\left(\sum_{i=1}^{n_p} ((q_{s,i} - q_{o,i})^2 q_{o,i}^2)\right)^{1/4}}{(\sum q_{o,i}^2)^{1/2}} \quad (3)$$

where  $r$  is the correlation coefficient between the simulated inflow and observed reservoir inflow,  $a$  and  $b$  are ratios of the standard deviation and mean of simulated inflows to the corresponding observed inflow respectively,  $q_s$  and  $q_o$  are the peak simulated and observed inflows respectively, and  $n_p$  is the number of peak flows greater than one-third of the mean peak flow observed. While KGE values closer to 1 indicate a better model performance, a PFC value closer to 0 signifies better peak flow simulation accuracy.

### 3.3. Bias-Correction

Each of the eleven ensemble precipitation forecasts from the GEFSv2 (Section 2.2.1) was bias-corrected by Empirical Quantile Mapping method [69]. The bias-correction of ensemble forecasts were performed using the reanalysis precipitation product of CaPA as a verifying database. Daily CaPA precipitation time series from January 2014 to December 2015 were used to bias-correct ensemble weather forecasts from January 2016 to December 2017 (Figure 3). That is, 4-years of daily ensemble precipitation forecasts were archived first (January 2014–December 2017). Then CaPA data were used as a training dataset for the first 2-years of ensemble forecasts. Parameters from the quantile mapping in the training period were applied to the last 2-years of ensemble forecast time series. This step is repeated for each ensemble member to produce a bias-corrected ensemble GEFSv2 inputs.



**Figure 3.** The Bias-correction process using the Empirical Quantile Mapping method.

### 3.4. Hindcast Simulation (Model Update and Forecast)

Hindcast simulation is performed in order to verify the hydrological models in forecast mode. The raw and bias-corrected Reforecast GEFS ensemble datasets were fed into the four calibrated hydrological models. The focus of the study is to assess the reservoir inflow forecast accuracy and skill of the models during the high flood periods. Therefore, 2017 is selected for forecast verification, which observes frequent spring and summer floods in the area. The hindcast period was from April 2017 to September 2017. Continuous model update and forecast were performed during the hindcast period (Figure 4). The hydrological models were run with observed meteorological data for at least one year before the forecast day in order to preserve and update the model's state parameters. In other words, the observed inputs were supplied to the models up to day-0. Then, ensemble forecasts were fed to the models for the next eight days, and this model update and forecast continuously every day during the entire hindcast period.

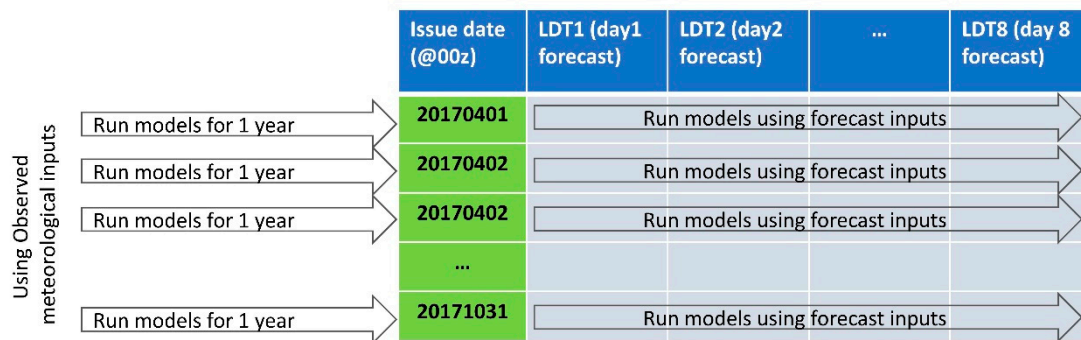


Figure 4. Model update and forecast in the hindcast period.

### 3.5. Ensemble Forecast Verification

Outputs from the previous step (model update and forecast) are daily ensemble reservoir inflow forecasts from four hydrological models for 1- up to 8-day lead times. The forecast skill and reliability of each model's ensemble reservoir inflow forecasts were evaluated using various ensemble verification metrics which are outlined below.

#### 3.5.1. Mean Continuous Rank Probability Score ( $\overline{CRPS}$ ) and Skill Score ( $\overline{CRPSS}$ )

The Mean  $CRPS$  measures the error of the commutative probability of the ensemble forecast. For an infinite number of classes or continuous variables,  $CRPS$  is calculated as follows [70]: Given cumulative distribution function of an ensemble  $y$  is  $P(y)$  and corresponding cumulative probability of observed value  $x$  with a step function  $1\{.\}$  representing 1 for ensemble values greater than observation and 0 otherwise, the  $CRPS$  and the mean  $CRPS$  can be computed by Equation (4).

The mean  $CRPS$  can be decomposed to mean reliability ( $\overline{Reli}$ ) and potential  $CRPS$  components, according to [71].  $\overline{Reli}$  is directly related to a rank histogram but provides more information. It measures the reliability of the system by examining whether the frequency of observations that falls in any one of ranked bins is equivalent to the other bins by taking into consideration the width of the bins, which the rank histograms do not do [71]. The potential  $CRPS$  ( $CRPS_{pot}$ ) is the  $CRPS$  of a perfect reliable system (i.e., when  $\overline{Reli} = 0$ ) or for a deterministic forecast where there is no spread.  $CRPS_{pot}$  is directly related to the spread of the ensembles and the presence of outliers [71]. The larger the spread or, the more outliers, the larger the  $CRPS_{pot}$ .  $\overline{CRPS}$ ,  $\overline{Reli}$  and  $CRPS_{pot}$  are negatively oriented, meaning a value of zero corresponds to a perfect ensemble forecast. Details of the derivation can be found on [71].

$$\begin{aligned}
 CRPS &= \int_{-\infty}^{\infty} (P(y) - 1\{y \geq x\})^2 dy \\
 \overline{CRPS} &= \frac{1}{n} \sum_{i=1}^n CRPS_i \\
 \overline{CRPS} &= \overline{Reli} + CRPS_{pot}
 \end{aligned}
 \quad (4)$$

Continuous Rank Probability Skill Score ( $\overline{CRPSS}$ ) is a scalar accuracy or performance measurement of the forecasting system by evaluating the mean continuous ranked probability score ( $\overline{CRPS}$ ) of ensembles with relative to a reference forecasting system [72]. It is positively oriented with a perfect score of 1 and is calculated by:

$$\overline{CRPSS} = 1 - \frac{\overline{CRPS}}{\overline{CRPS}_{ref}} \quad (5)$$

For the reference forecasting system, we used the climatological ensembles of the last twenty-four years of historical daily reservoir inflows. This is practically used by Manitoba Hydrological Forecasting Center to issue medium- and long-term ensemble forecasts at the site [73]. Reference [74]



discussed the option of using climatological observations as an alternative benchmark hydrological ensemble prediction.

### 3.5.2. Reliability Diagram

The reliability diagram also called Attribute Diagram by Hsu et al. [75], is a measure of the accuracy of ensemble forecasts, which plots the observed relative frequency with respect to forecasting probability in different bins of the category [70]. It is a plot of forecast probability versus observed frequency, and perfect reliability is indicated by a curve lying along the diagonal of a reliability diagram [76].

The CRPS decomposition parameters of Hersbach et al. [71], were used to draw reliability diagrams in this study. There are 90% confidence intervals applied for the reliability diagrams using the bootstrap resampling technique to measure the conditional verification pair sample uncertainty.

### 3.5.3. Relative Operating Characteristics (ROC) and Skill Score (ROC Score)

ROC is a powerful metric to measure the probabilistic forecast occurrence of events across a range of thresholds [77]. For each threshold, ROC examines the correspondence between the forecast and observation by defining the probability of detection (hit rate) and the probability of false detection (False alarm rate). ROC curve for several thresholds can then be constructed by 'Hit Rate' values as ordinate and 'False Alarm Rate' values as abscissa. A good and skillful forecast produces a ROC curve above the 45 degrees diagonal, but more towards the top-left position, indicating high 'Hit Rate' and low 'False Alarm Rate' [77]. ROC shows the discrimination skill of the ensemble forecast system [78,79]. Discrimination skill indicates the ability of the forecasting system to categorize occurrence and non-occurrence of floods defined between user-defined probability thresholds [78–80].

A single scalar score can summarize the quality of ROC curves. ROC score is a function of the area under the ROC curve (AUC). Wilks [70], formulates a simple equation for ROC Score as:

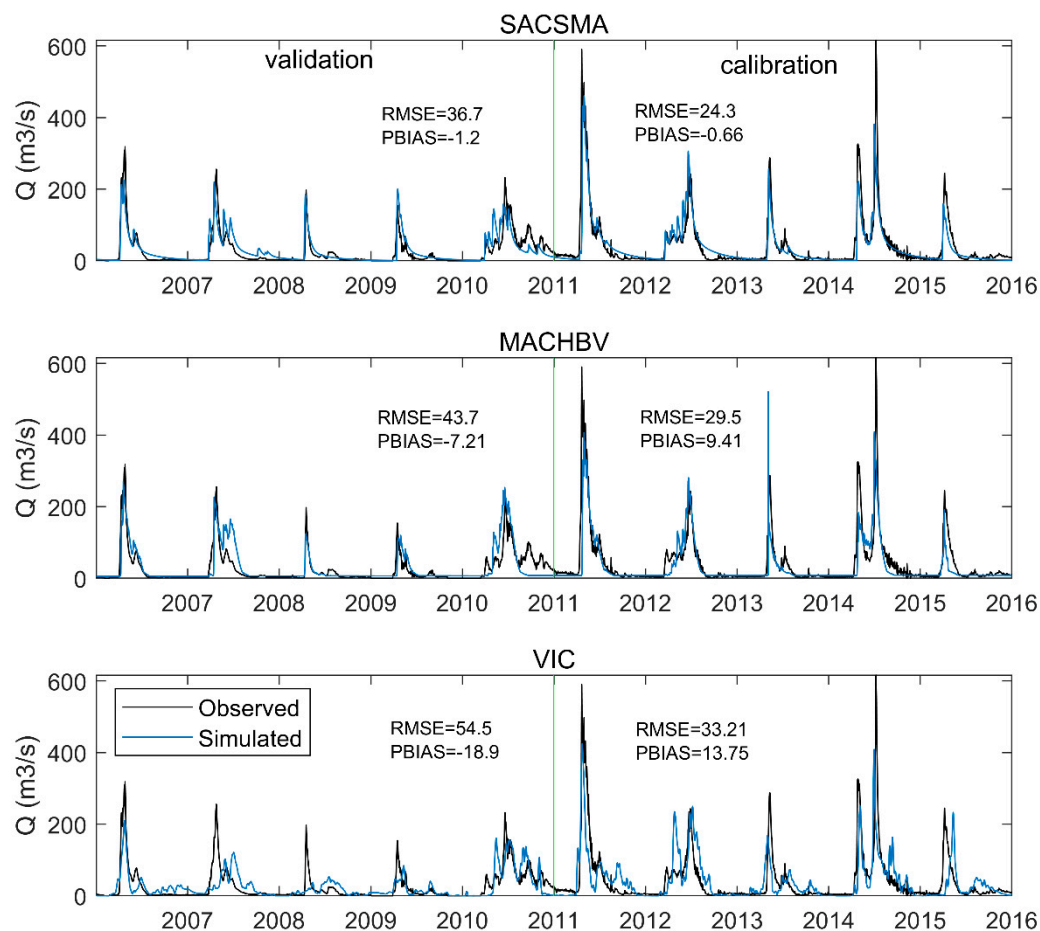
$$ROC\ Score = 2 * AUC - 1 \quad (6)$$

where AUC is the area under the curve of each Relative Operating Characteristics curves. A perfect system that has ROC curves close to the top left corner would have a score of 1.

## 4. Results

### 4.1. Calibration and Validation

As noted in Section 3.2, DDS optimization was used to calibrate parameters of SACSMA with SNOW17, MACHBV with SNOW17, and VIC with RVIC models. Among the ten years chosen for calibration and validation, the recent five consecutive years (from 2011 to 2015) were used to calibrate the models, and the previous five years (2006 to 2010) were used to validate the models (Figure 5). The reason why we used recent data for calibration is that we want to train the models using high consecutive flood periods. Looking at the historical time series from 2006 to 2015, the recent five-years are high consecutive flood years than the previous five-years. Moreover, it is highly likely that this trend will continue past 2016 and the near future due to anticipated climate change impact in the region and other similar factors that caused the recent high consecutive flood years. Since the challenge of achieving the accurate reservoir inflow forecasting arises particularly during flood periods, and the objective of the paper focuses on improving the accuracy of flood forecasting in large complex watersheds, the hydrological models were trained/calibrated with the recent flood years.



**Figure 5.** Calibration and validation plots of SACSMA with SNOW17, MACHBV with SNOW17, and Variable Infiltration Capacity (VIC) with routing module (RVIC) models. RMSE: Root-Mean-Square Error, and PBIAS: Percent Bias [81] are displayed for each model to provide more information in addition to the visual inspection of the time series.

The performance metrics of the models are summarized in Table 1. The KGE performance statistics indicate that the SACSMA model outperforms MACHBV followed by VIC during calibration as well as validation periods. The lumped models (SACSMA and MACHBV) appear to show better performances than the macroscale model (VIC). The Peak Flow Criteria (PFC) shows that SACSMA and MACHBV have improved and have comparable accuracy in peak flow prediction. The VIC model slightly underestimates and delays peak flows occasionally, although it maintains the hydrograph during spring and summer high inflow seasons.

**Table 1.** Performance statistics of the three hydrological models from calibration and validation. The definition of the abbreviations is presented in Section 3.2.

	Calibration			Validation		
	SACSMA	MACHBV	VIC	SACSMA	MACHBV	VIC
PFC	0.180	0.174	0.247	0.234	0.231	0.270
KGE	0.796	0.740	0.653	0.776	0.679	0.653

The performance of the models can also be seen from the simulated and observed flow hydrographs shown in Figure 5. Visual inspection shows that all three models comparatively capture the pattern of the observed reservoir inflow hydrographs during calibration and validation periods; although SACSMA and MACHBV models appear to reproduce the peak flows better than VIC. In addition to

the visual inspection, the RMSE and PBIAS values of each model are displayed in Figure 5 to provide more information on the hydrographs. It can be seen that SACSMA provided better accuracy and less bias during calibration and validation, followed by MACHBV and VIC models in decreasing order of performances.

The calibration of the models was performed in a daily time step, and the optimization method (DDS algorithm) used during the calibration was the same for all models. The objective function is also the same, which is the average of *KGE* and *PFC*. However, there are differences in the number of parameters (dimensionality) among the models. Note that the lumped models were coupled with SNOW17, hence the total number of the calibrated parameters are the summation of individual models' parameters; for example, SACSMA (15) plus SNOW17 (10). As noted in Appendix A.4, the default number of VIC/RVIC model parameters was further refined to improve the calibration output and to better represent the wetland, landcover, and soil types of the basin. The parameters were refined and increased from the default 13 to 53 based on land cover classes and soil mapping units (Figure A1). A simple test has been done before refining the parameters by performing the calibration using the default 13 parameters, and the preliminary results were much worse ( $KGE = -2.3$ , not shown in the results section) than after refining the parameters ( $KGE = 0.653$ ). With the default parameters, VIC, as a macroscale land-surface based model, was not correctly estimating the water and energy balance equations in a vertical column at each grid cell and transferring water between grids and river networks by using the routing module (RVIC). After refining, the model significantly improved the water interaction in wetland areas, and different land cover and soil tiles and routed the flow to the outlet.

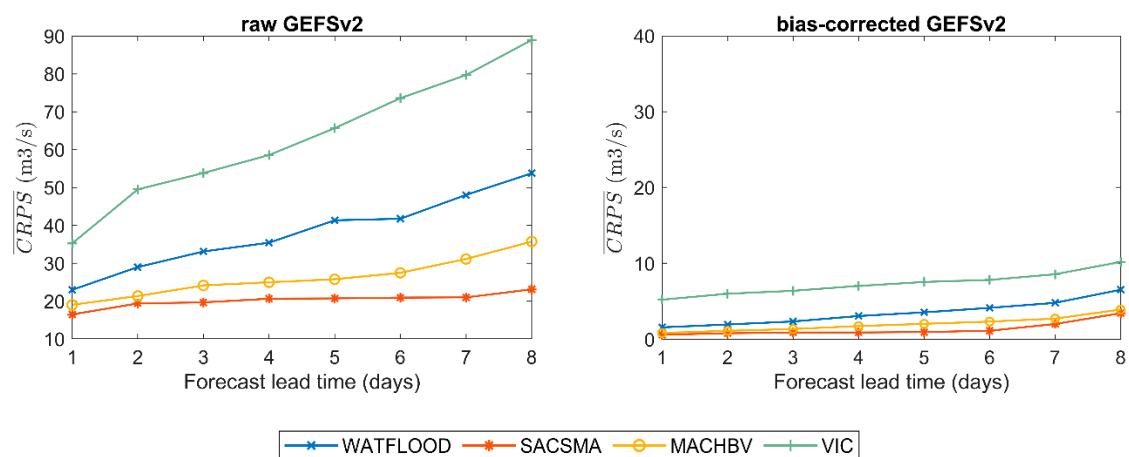
The message here is that an effort has been made to employ a better calibration approach with an efficient optimization algorithm for the advanced model (VIC). As discussed in Section 3.2, DDS is a competitive and efficient optimization tool that has been applied in several distributed and land-surface based hydrological models. Thus, it is safe to say that the conclusion (i.e., the improved performance of SACSMA and MAHBV over VIC in the calibration outputs) was not limited by the search algorithm.

## 4.2. Model Comparison in Forecast Mode

### 4.2.1. Overall Forecast Quality and Skill

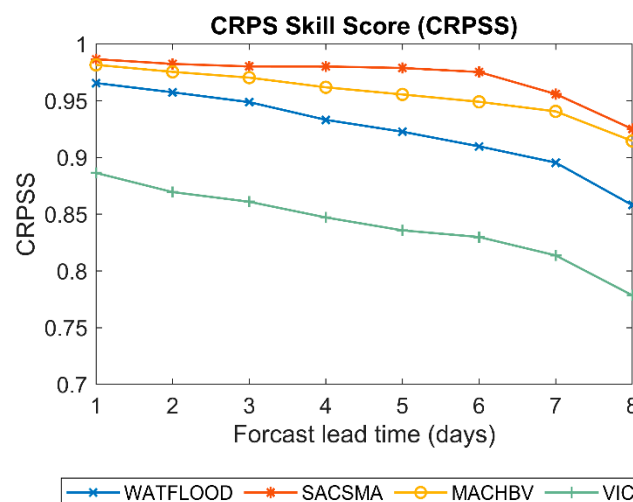
GEFSv2 ensemble precipitation forecasts were bias-corrected by the Empirical Quantile Mapping method using CaPA as a verifying analysis, as described in Section 3.3. Both raw and bias-corrected GEFSv2 inputs were fed into four hydrological models in order to (1) realize the effect of the bias-correction on the output hydrological forecasts, and (2) compare the models forecast performance pre- and post-bias correction process.

Figure 6 shows the mean *CRPS*, which measures the overall probabilistic error of the ensemble reservoir inflow forecasts generated by four hydrological models and GEFSv2 inputs. As expected, the bias-corrected GEFSv2 ensembles significantly outperform the raw GEFSv2 inputs regardless of the hydrological models used. The quality of hydrological forecasts was much improved by bias-correcting each ensemble precipitation forecast of GEFSv2 with CaPA reanalysis data. Figure 6 also shows a comparison between the forecast quality of the four hydrological models. For all models, the overall forecast quality declines as the lead time increases, as expected. It can be seen from the figure that the mean *CRPS* values of the SACSMA model are the lowest followed by MACHBV, WATFLOOD, and VIC in ascending order of forecast probability error. Whether using raw GEFSs or bias-corrected GEFSs as in inputs, the resultant hydrological forecast skill of the two lumped models (SACSMA and MACHBV) outperforms the benchmark distributed WATFLOOD model and the macroscale VIC model at all lead times. However, the benchmark model provides a better skill than VIC and is relatively close to the two lumped models at early lead times.



**Figure 6.** Mean CRPS of ensemble reservoir inflows generated with Raw (left) and Bias-corrected (right) ensemble GEFSv2 precipitation forecasts.

So far, the models' ensemble outputs were evaluated based on their overall forecast error. In order to add a comprehensive outlook, a reference ensemble forecasting system is used to evaluate their skills. Figure 7 shows the mean CRPS skill score (CRPSS) of ensemble reservoir inflows simulated by the four different hydrological models. It can be seen from the figure that, the CRPSS values of the four models have a similar trend, as the CRPS depicting that the lumped models have better forecast skill than the benchmark and macroscale models. Comparing to the reference climatological-based ensembles, SACSMA provides the best quality of ensembles at all lead times, followed by MACHBV, WATFLOOD, and VIC. The lumped models were competitive throughout the forecast horizon, with minor exceptions. For the first two to three days, the skill score of the benchmark WATFLOOD is relatively close to the lumped models, but the forecast skill gradually deteriorates at later lead times.



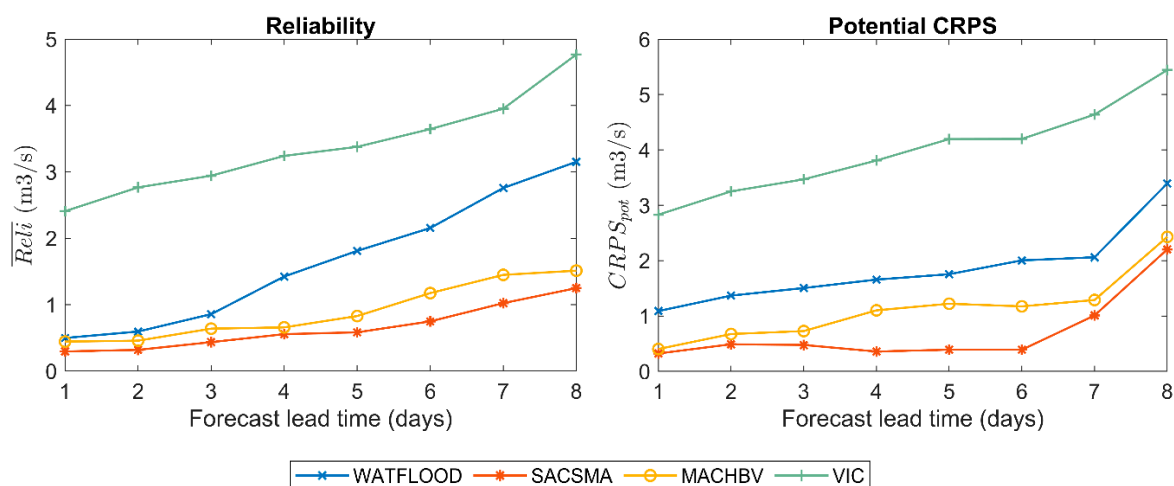
**Figure 7.** Comparison of reservoir inflow ensembles between four hydrological models using CRPS skill score (CRPSS).

#### 4.2.2. Reliability

The reliability of the ensemble hydrological forecasts was evaluated by two metrics; using the reliability component of CRPS after decomposition of [71], and using the Reliability Diagram.

Figure 8 shows the components of CRPS after decomposition of [71]. Here, the summation of the reliability (left) and potential CRPS (right) components of each hydrological model is the mean CRPS. The reliability and potential CRPS components follow the same trend as the mean CRPS and

CRPSS. The decomposition indicates that the lumped models (SACSMA and MACHBV) were more reliable and have less spread and outliers than the benchmark WATFLOOD and macroscale VIC models. The reliability component contributes about half of the mean CRPS. The rest comes from the potential CRPS. Remarkably, it can be observed that the forecast quality of WATFLOOD during the first two or three days comes from the reliability component because this value is lower and much closer to the lumped models than the potential CRPS component. The overall forecast quality of the SACSMA model remarkably remains the same up to lead time of day 6, as can be seen from mean CRPS and CRPSS values. This effect is mainly due to the potential CRPS component, which remains either constant or slightly dropped as going from day 1 to day 6. SACSMA model generates ensembles that are less spread and have low number of outliers in the first six days forecast as explained by the potential CRPS component. The potential CRPS rapidly increased in all models after lead time seven, which indicates that the ensemble spread and presence of outliers start to significantly rise irrespective of the model type after a seven-day forecast. The sudden rise and decline of mean CRPS and CRPSS in most models after day seven, maybe due to this effect.

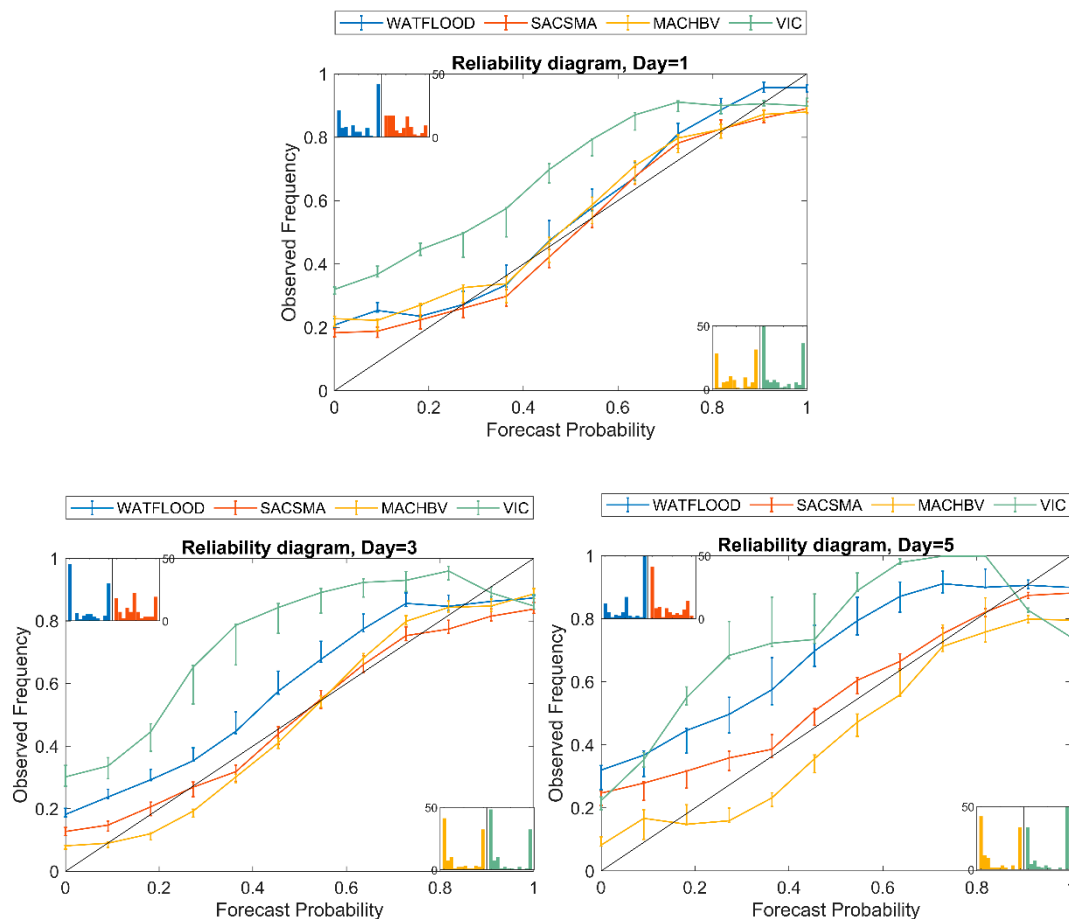


**Figure 8.** CRPS decomposition components. The left plot shows the reliability component, and the right plot shows the Potential CRPS component. A comparison of these attributes was made between four hydrological models.

Figure 9 shows the reliability diagrams of the hydrological models for forecast lead time of day 1, 3, and 5. For a one-day lead time forecast, the reliability curves of SACSMA, MACHBV, and WATFLOOD were all reasonably aligned along the diagonal line, which indicates that they achieve relatively more reliable forecasts. The conditional observed frequency is comparable with the forecast probability with slight exceptions in the very lower bin. For day three forecasts, this trend minimally changes, but overall, the reliability of the WATFLOOD is not significantly lower than the lumped models. For day five, the reliability curve of SACSMA is still close to the diagonal ('perfect line'), especially on higher forecast probabilities. MACHBV is relatively reliable on day five of the forecast, as shown by its diagram. However, the reliability curve of WATFLOOD at day five is away from the diagonal line indicating its reliability was progressively declining after day three forecast. The reliability of VIC, although relatively moderate at day one, was reduced at day three and five forecasts because it continuously underestimates the forecast. The 90% Confidence Intervals (CI) of the reliability diagrams showed that uncertainties in the conditional verification pair samples increased in all models as the forecast lead time increases. However, the advancement of conditional uncertainty in the forecasts in lumped models was not substantial when compared to the benchmark and macroscale models. This is because the reliability lines of SACSMA and MACHBV are within the CI bounds most of the time, and their CI's are closer to the diagonal line. Whereas, for VIC, the reliability lines are either at the lower or upper level of the CI's in all cases and for WATFLOOD this occurs on day three lead time. This



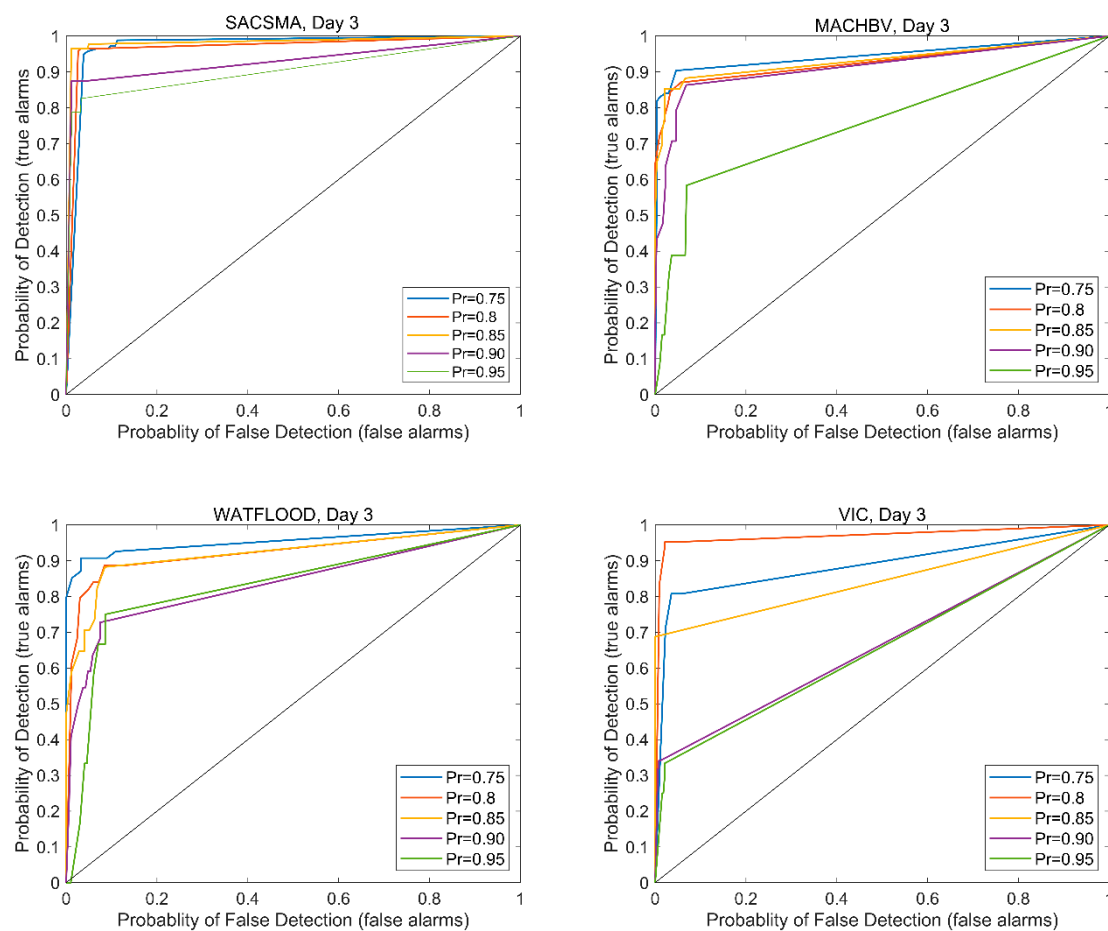
characteristic indicates that 90% of the cases, the reliability diagram attributes of VIC, and sometimes WATFLOOD did not belong to the interval where the “true” value of the attributes exists, whereas for the lumped models this does not hold.



**Figure 9.** Reliability diagrams of ensemble reservoir inflows at three selected forecast lead times. Different colors show different model types. The 90% confidence intervals were shown in the reliability lines for each model. The inset histograms show the frequency of occurrence in each forecast bin.

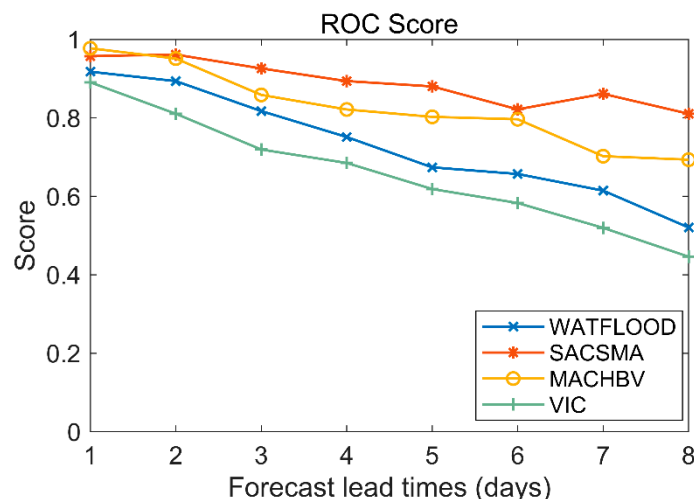
#### 4.2.3. Hit and False Alarm Rate Distribution

As described in Section 3.5.3, ROC displays the hit-rate and false alarm rate of a forecasting system at different thresholds. Figure 10 shows the ROC curves of the models at day-3 forecast lead time. The hit rate versus false alarm rates was drawn for varying higher probability threshold levels of reservoir inflows because the primary focus of this research is on flood forecasting. Simulated ensemble inflows exceeding 75, 80, 85, 90, and 95 percentiles of the observed reservoir inflow were taken into consideration. At day three lead time (Figure 10), SACSMA performed well in attaining the highest true alarm and lowest false alarm rates for all probability thresholds, as compared to other models, the closest one being MACHBV. Forecasting the most extreme flood or flows exceeding 95 and 90 percentile inflows is a challenge that most models lack with different levels of forecast skill. The lumped models and subtly of the benchmark are deemed sufficient to construct ensembles that have good discrimination skills to forecast up to 85 percentile reservoir inflow. Although the ROC curves stipulate that VIC can, in fact, reproduce 80 percentile flows up to five days ahead forecast time other probability thresholds have almost zero skills.



**Figure 10.** Relative Operating Characteristic (ROC) curves drawn for probability thresholds exceeding 75, 80, 85, 90, and 95 percentile reservoir inflows for three days ahead forecast. The four plots are for four different hydrological models.

In Figure 11, the ROC Scores of each model, estimated by the average of the area under the ROC curves for the considered probability thresholds, are shown. It summarizes the performance and discrimination skills of the models' ensembles for all forecast time horizons. The ROC Scores indicate that the forecast skills of WATFLOOD and VIC monotonically decrease as the lead time increases, but for the case of SACSMA and MACHBV, even though their skill unevenly decline, they have competitive and relatively decent forecast performances. The declining ROC Scores indicate that as the lead time increases, the ROC curves (not shown here) progressively approach the diagonal line, which is the climatological forecast or "zero skill" line [78]. In general, considering all the forecast lead times and probability thresholds, SACSMA appeared to have better discrimination skills more than the others, followed by MACHBV, WATFLOOD, and VIC in order of performance.



**Figure 11.** The ROC score measured by the Area Under the ROC Curves of four hydrological models.

## 5. Conclusions and Discussion

The objective of this study was to identify hydrological models from a pool of diverse model structures that can produce better forecast skill and reliability and provide an enhanced short- and medium-range reservoir inflow forecasts in a Prairie watershed: The Upper Assiniboine River Basin. A comparison of forecast skill and reliability between the selected hydrological models was made using raw and bias-corrected ensemble precipitation forecast products. The best model was selected from two lumped models (SACSMA with SNOW17 and MACHBV with SNOW17), a benchmark distributed model (WATFLOOD), and macroscale land-surface based model (VIC). Daily total precipitation forecasts were collected from an 11-member second-generation Global Ensemble Forecast System reforecast (GEFSv2). Each of the ensemble members was bias-corrected by Empirical Quantile Mapping method using the Canadian Precipitation Analysis (CaPA), as a training/verification dataset. Raw and bias-corrected GEFSv2 precipitation were supplied to the hydrological models to evaluate and compare the forecast skill and reliability of the ensemble inflow outputs. Forecast evaluation was performed in a 6-month hindcast period where daily ensemble reservoir inflow forecasts were issued for 1-day up to 8-days forecast lead times. SACSMA, MACHBV, and VIC models were calibrated in the study area by comparing simulated and observed inflows into Shelmouth Reservoir while the WATFLOOD model, which is operationally implemented for the Provincial real-time flood forecasting was used as a benchmark.

Results indicated that simulated ensemble reservoir inflows generated by bias-corrected GEFSv2 provided significantly better forecast quality than the raw GEFSv. Even though this result is expected, two things are noticed; first, the bias-correction of each ensemble members instead of the mean or median provided a consistent and reliable ensemble inflow forecast, and second, bias-correcting forecasts using verifying datasets (such as CaPA) for a training period of at least two years before the forecast time results in an improved hydrological forecast. This method and the improved result can be beneficial for users at operational flood forecasting centers, as they would generally prefer less advanced and quick post-processing methods.

All models were supplied with bias-corrected ensemble GEFSv2, and various ensemble verification metrics were used to compare the model outputs up to eight days of forecast lead times. The overall forecast quality and skill of the models' results were evaluated by using mean CRPS and CRPSS metrics. Results indicated that the two lumped models (SACSMA and MACHBV) provided better overall forecast performance than the benchmark WATFLOOD and the macroscale VIC models. Although the lumped models (SACSMA and MACHBV) were found to be comparable, SACSMA provided enhanced forecast skill than MACHBV at all lead-times. For shorter lead-times, particularly up to day 3, WATFLOOD provided relatively competitive overall forecast quality as of the lumped models.

The *CRPS* decomposition by [71] was found to be vital to interpret and better analyze the overall forecast performance. This decomposition indicated that the modest forecast skill of WATFLOOD in the first 2 or 3 days came from the reliability component of *CRPS*. The decomposition of *CRPS* further indicated that the superior performance of SACSMA is due to its ability to generate ensemble inflow forecasts with less ensemble spread and low presence of outliers in the first six days of the forecast as explained by the potential *CRPS* component.

Reliability diagrams of the hydrological models at different lead times provided further insight into the forecast skill of the ensembles. At shorter lead times, the reliability diagrams of SACSMA, MACHBV, and WATFLOOD indicated that they all achieve relatively reliable forecasts as the conditional observed frequency was comparable with the forecast probability. However, after day 5, the reliability of WATFLOOD deteriorated while MACHBV and SACSMA, in order of increasing performance, remain within reasonable calibration accuracy. The reliability of VIC, although relatively moderate at day one, was weak because it continuously underestimated the forecast.

In order to evaluate the discrimination skill of the ensembles, two threshold-based metrics were used to evaluate the hit-rates and false-alarm rates at different higher forecast thresholds: Relative Operating Characteristic (ROC) curve and the ROC Score measured by the area under the ROC Curves. ROC curves of the models were drawn and compared for ensemble reservoir inflows exceeding between 75 and 95 percentiles, with a 5 percent increment. For day three forecast, SACSMA and MACHBV models attained highest true alarm, and lowest false alarm rates for all probability thresholds with the former slightly outperformed the later. As the lead time increases, forecasting the most extreme flows exceeding 95 percentile inflows was a challenge for most models. However, the lumped models and moderately the benchmark were sufficiently able to generate ensemble inflows that have very good skills to forecast inflows exceeding the 85 percentiles. Overall, considering all the forecast lead times and probability thresholds, SACSMA provided better differentiation skill than the others, followed by MACHBV, WATFLOOD, and VIC in order of decreasing performance.

In general, ensemble inflow forecasts generated by the lumped models offered substantially better performances as compared to the benchmark distributed model or the macro-scale land surface models. The distributed benchmark model unequivocally provided reliability as good as the lumped models up to three days ahead even though it deteriorates rapidly at later lead times. It is anticipated that the forecast performance of the VIC model could be improved by increasing the grid resolution of the model, which was set up at 1/8-degree horizontal resolution. Overall the SACSMA appeared to generate the most reliable and skillful ensemble reservoir forecast inflows for up to a week ahead lead times and should be considered as an alternative operational model in the study area.

The performance of different hydrological models depends on many factors such as the scale, the complexity of the basin, the spatial and temporal resolution of the input data, the structure of the models, the degree of discretization of the models, and the number of parameters to be calibrated, etc. For the models that were applied to this research, these factors are interconnected and thus affect their calibration performance jointly. The intended purpose of the hydrological models in this study is to simulate and forecast short- and medium-range reservoir inflows. Regardless of the structure and degree of discretization of the models, the objective is to obtain a time series (hydrograph) implicitly at one location, which is considered as the watershed outlet, and no interior locations or sites are needed. The way the inputs were supplied to the models depends on the type of the model (e.g., lumped, distributed) and the discretization level (e.g., spatially lumped catchment, grids, GRUs, HRUs). Hence, it can be said that the calibration performance of the hydrological models was influenced jointly by the above factors.

Moreover, there are many references from the literature where lumped models outperformed various distributed or land-surface based models. The Distributed Model Inter-comparison Project (DMIP) has implemented several hydrological models at eight basins of the River Forecasting Centres in the USA, and the results showed that lumped models (particularly SACSMA) provided better performance than distributed (such as WATFLOOD, SWAT) and land-surface based models (such

as VIC, NOAH) [82]. Results from the Model Parameter Estimation Experiment (MOPEX) also demonstrated a significantly improved performance of SACSMA comparing to land-surface based models including VIC and SWAP [83,84]. Maurer et al. [85], performed a comparative study between SACSMA and VIC models, and the results revealed that the former lumped model had an evident better calibration performance over the later land-surface model.

The research is conducted to identify the best performing hydrological models for improved hydrological forecasting in a specific large complex watershed of the prairie region of Canada. The Upper Assiniboine Basin is characterized as a “Prairie” watershed, which is known for its complex hydrology due to the presence of potholes. Hence, the study area is considered as one example of a complex watershed. The hydrological models have a diversified structure (lumped, distributed, and macro-scale land-surface based) and implemented to evaluate and select the model that has the best potential for simulating and predicting reservoir inflows for such a complex basin. If another kind of complex watershed with the same scale is used, it is believed that a similar conclusion would be drawn. The previous studies that provided similar conclusions were conducted in various watershed landscapes. The MOPEX project was tested in twelve watersheds that have various land cover types such as croplands, mixed forests and natural vegetation in different altitudes [83,84]. In the DMIP study, the dominant land cover properties of the eight basins were mainly agriculture and forests with varying topographies and soil types [82,86]. The comparative study of Maurer et al. [85] was performed in snow-dominated catchments. Overall, the same candidate model(s) would highly likely be identified to better simulate and forecast medium-range reservoir inflows in other types of complex watersheds with a similar scale and characteristics.

In general, for hydrological forecasting focusing on basin outflows and not interior sites, the study indicated that lumped models, particularly SACSMA with SNOW17, provided better performance than the distributed or land-surface models in complex watersheds. Not only the calibration but also the validation and forecast verification analysis have given the superiority in simple models. The verification of hydrological forecasts generated from bias-corrected ensemble weather forecast inputs provided enough details of the model’s performances for the intended purpose.

**Author Contributions:** Conceptualization, F.S.A. and P.C.; Data curation, F.S.A.; Formal analysis, F.S.A.; Investigation, F.S.A.; Methodology, F.S.A. and P.C.; Software, F.S.A., P.C., and F.U.; Supervision, P.C. and I.T.; Validation, F.S.A. and P.C.; Resources, P.C., I.T., and F.U.; Writing—original draft preparation, F.S.A.; Writing—review and editing, P.C. and F.U.

**Funding:** This work was supported by the Natural Science and Engineering Research Council (NSERC) Canadian FloodNet (Grant number: NETGP 451456).

**Acknowledgments:** The authors would like to thank the Manitoba Infrastructure, Hydrologic Forecasting Center for providing the operational WATFLOOD model along with data.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A Brief Description of the Calibration of Models

### Appendix A.1 SNOW17

The Snow Accumulation and Ablation Model (SNOW17) model was developed by [44] as part of the NWS river forecasting system. It is a conceptual model that uses a temperature index to determine energy exchange across snow-air interface [44].

Inputs to the model are:

- (i) The mean area observed precipitation time series obtained by Thiessen Polygon method
- (ii) The mean area observed temperature time series obtained by Thiessen Polygon method
- (iii) The average elevation of the catchment
- (iv) The latitude of the centroid of the catchment

The parameters that were calibrated are listed in Table A1.



The MATLAB version of the source code was used to set up and calibrate the model in the study area.

**Table A1.** SNOW17 model parameters.

No	Parameters	Description	Unit	Ranges
1	SCF	Snowfall correction factor	–	0.4–1.6
2	MFMAX	Maximum melt factor during non-rain periods considered to occur on June 21	mm/6 h/°C	0.5–2.0
3	MFMIN	Minimum melt factor during non-rain periods considered to occur on December 21	mm/6 h/°C	0.05–0.5
4	UADJ	The average wind function during rain-on-snow periods	mm/mb/°C	0.03–0.2
5	NMF	Maximum negative melt factor	mm/6 h/°C	0.05–0.50
6	MBASE	Base temperature for non-rain melt factor above which melt typically occurs	°C	0–2.0
7	PXTEMP1	Lower Limit Temperature dividing transition from snow, if temp is less than or equal to pxtemp1, all precip is snow. Otherwise it is mixed linearly	°C	–2.0 to 0
7	PXTEMP2	Upper Limit Temperature dividing transition from snow, if temp is greater than or equal to pxtemp2, all precip is rain. Otherwise it is mixed linearly	°C	1 to 3.0
8	PLWHC	percent liquid water holding capacity of the snow pack	–	0.02–0.3
9	DAYGM	Daily melt at snow–soil interface	mm/day	0–0.3
10	TIPM	Antecedent snow temperature index	–	0.1–0.2

Outputs from SNOW17 are outflow and Snow Water Equivalent. The outflows are the summation of snowmelt and rain. The coupling mechanism of SNOW17 with SACSMA and MACHBV models is performed by forcing outflows from SNOW17 into the hydrological models.

#### Appendix A.2 SACSMA

The Sacramento Soil Moisture Accounting (SAC-SMA) model has been used as a lumped conceptual model at the National Weather Service (NWS) for operational river forecasting purposes. It has also been included within the National Weather Service Hydrology Laboratory's Research Distributed Hydrologic Model (HL-RDHM) by adding several processes [87]. Details description of the lumped SACSMA model can be found in [88].

In this research, SACSMA was implemented as a lumped continuous model in Upper Assiniboine Basin. The MATLAB version of the source code was used to set up and calibrate the model.

For calibrating the model, the following inputs are used:

- (i) Outflow (rain plus snowmelt) from SNOW17
- (ii) The catchment area of the basin
- (iii) Observed catchment outflow estimated by calculated reservoir inflow

The lists the parameters of the model that were calibrated by DDS optimization are presented in Table A2.

**Table A2.** SACSMA model parameters.

No	Parameters	Description	Unit	Ranges
1	UZWWM	Upper zone tension water maximum storage	[mm]	1–150
2	UZFWM	Upper zone free water maximum storage	[mm]	1–150
3	LZWWM	Lower zone tension water maximum storage	[mm]	1–500
4	LZFPM	Lower zone free water primary maximum storage	[mm]	1–1000
5	LZFMS	Lower zone free water supplemental maximum storage	[mm]	1–1000
6	ADIMP	Additional impervious area	[-]	0.0–0.4
7	UZK	Upper zone free water lateral depletion rate	[day <sup>-1</sup> ]	0.1–0.5
8	LZPK	Lower zone primary free water depletion rate	[day <sup>-1</sup> ]	0.0001–0.025
9	LZSK	Lower zone supplemental free water depletion rate	[day <sup>-1</sup> ]	0.01–0.25
10	ZPERC	Maximum percolation rate	[-]	1–250
11	REXP	Exponent of the percolation equation [-]	[-]	1–5.0
12	PCTIM	Impervious fraction of the watershed area	[-]	0.0–0.1
13	PFREE	fraction percolating from upper to lower zone free water Storage	[-]	0.0–0.6
14	athorn	A constant for Thornthwaite's equation	[-]	0.1–0.3
15	Rq	Routing Coefficient	[-]	0.0–1

### Appendix A.3 MACHBV

McMaster University-Hydrologiska Byråns Vattenbalansavdelning (MAC-HBV) is a modified version of the lumped conceptual HBV model edited by [45] at McMaster University. Detail description of the model can be found in [89]. MACHBV has been implemented in several Canadian watersheds for flood forecasting purposes [90–94].

The model was calibrated in Upper Assiniboine River Basin in this study. The MATLAB version of the source code was used to set up and calibrate the model. The following inputs were used for calibration:

- (i) Outflow (rain plus snowmelt) from SNOW17
- (ii) The catchment area of the basin
- (iii) Observed catchment outflow estimated by calculated reservoir inflow

The lists the parameters of the model that were calibrated by DDS optimization are presented in Table A3.

**Table A3.** MACHBV model parameters.

No	Parameters	Description	Unit	Ranges
1	athorn	A constant for Thornthwaite's equation	[-]	0.1–0.3
2	fc	Maximum soil box water content	[mm]	50–800
3	lp	Limit for potential evaporation	[mm/mm]	$0.1 \times fc - 0.9 \times fc$
4	beta	A non-linear parameter controlling runoff generation	[-]	1–10
5	K0	Flow recession coefficient in an upper soil reservoir	[days]	1–30
6	lsuz	A threshold value used to control response routing on an upper soil reservoir	[mm]	1–100
7	K1	Flow recession coefficient in an upper soil reservoir	[days]	2.5–100
8	cperc	A constant percolation rate parameter	[mm/day]	0.01–6
9	K2	Flow recession coefficient in a lower soil reservoir	[days]	20–1000
10	maxbas	A triangle weighting function for modelling a channel routing routine	[days]	1–20
11	rcr	Rainfall correction factor	[-]	0.5–1.5
12	a1	An exponent in relation between outflow and storage representing non-linearity of storage – discharge relationship of lower reservoir	[-]	0.5–20

#### Appendix A.4 VIC

The Variable Infiltration Capacity (VIC) model is a Macroscopic Land-surface distributed hydrological model. Detail description of the model can be found in [46]. Since VIC computes its energy and water balance equations in a vertical column at each grid cells, an external river routing module is required to route runoff and baseflows to the edge of each grid cell throughout the river network to the catchment outflow [59]. For this purpose, the python version of RVIC routing module [58] is used in this research. Version 4.2.d of VIC was setup and coupled with RVIC.

Meteorological forcing to the model are:

- (i) Average daily gridded interpolated precipitation data from the ground network,
- (ii) Daily gridded minimum, and maximum temperature data from ANUSPLIN,
- (iii) Average daily wind speed from the Global Environmental Multiscale (GEM) model.

Physical inputs are:

- (i) Digital elevation model for SNOW elevation bands and flow direction computation
- (ii) Land cover data from Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type (MCD12Q1) Version 6 data product
- (iii) Soil data from FAO's Harmonized World Soil Database (HWSD) V 1.2

The grid resolution of VIC model setup was about 1/8 degree. At each grid, three elevation bands and three soil layers were used in the study area.

Some of the main processes are described below:

- (i) Snow: Rain-snow partitioning, snow accumulation, and melting are simulated at a sub-grid level using temperature index method lapsed through the Elevation (SNOW) bands.
- (ii) Evaporation: is simulated at each elevation band and land cover type using Penman-Monteith Approach.

For this study, the dynamic wetland module was activated to calibrate the wetland parameters because as a Prairie watershed, the area has abundant wetland and potholes, as shown in Figure A1. The default parameters of the VIC/RVIC model (Table A4) were further refined to improve the calibration output and to better represent the physical characteristics of such a large complex basin. Wetland parameters were refined based on the vegetation type in the catchment. Each of the three major land cover classes (Figure A1) in the area has been assigned with its own five wetland parameters. Similarly, the six soil parameters were sub-categorized into six soil groups based on the dominant and associated soil types of the basin (Figure A1). After parameter refining, the total number of parameters to be calibrated was increased from the default 13 to 53 including the routing parameters.

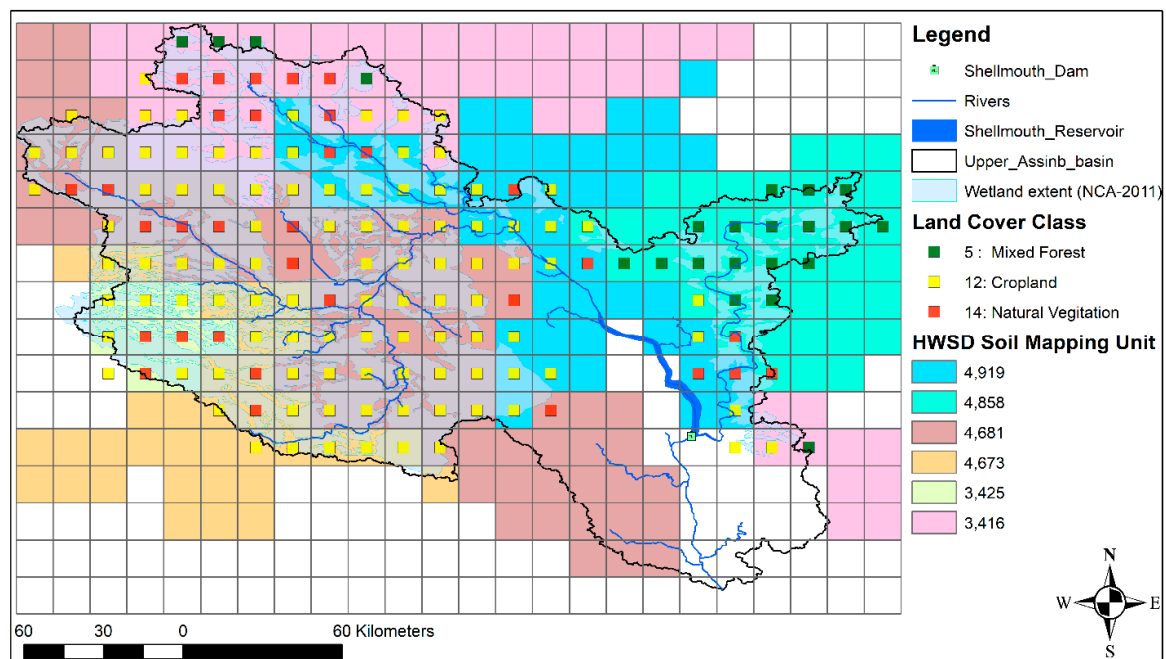


Figure A1. Soil and Land cover tiles discretization for VIC model.

Table A4. VIC/RVIC model parameters.

No	Notation	Range	Unit	Definition
<b>Soil parameters</b>				
1	b	$10^{-5}$ –0.4	-	Variable infiltration curve parameter
2	Ds	$10^{-3}$ –1	-	Fraction of Dsmax where non-linear baseflow begins
3	Dm	0.1–30	mm/day	Maximum velocity of baseflow
4	Ws	0.5–1	-	Fraction of maximum soil moisture where non-linear baseflow occurs
5	s2	0.3–1.5	m	Thickness of middle soil moisture layer
6	s3	0.3–1.5	m	Thickness of bottom soil moisture layer
<b>Wetland parameters</b>				
7	bmin_depth	0.01–0.3	m	Lake depth below which channel outflow is 0.
8	wfrac	0.001–0.05	-	Width of lake outlet, as a fraction of the lake perimeter
9	depth_in	0.01–0.3	m	Initial lake depth
10	rpercent	0.1–1	-	Fraction of grid cell runoff that enters lake (instead of going directly to channel network)
11	lake_depth	0.1–1.5	m	Maximum allowable depth of lake
<b>Routing parameters</b>				
12	Vl	0.5–3	m/s	Flow/Wave velocity
13	Df	200–4000	m <sup>2</sup> /s	Flow diffusion

## References

1. Fang, X.; Minke, A.; Pomeroy, J.; Brown, T.; Westbrook, C.; Guo, X.; Guangul, S. *A Review of Canadian Prairie Hydrology: Principles, Modelling and Response to Land Use and Drainage Change*, Centre for Hydrology Report #2, Version 2; University of Saskatchewan: Saskatoon, SK, Canada, 2007.
2. Armstrong, R.N.; Pomeroy, J.W.; Martz, L.W. Evaluation of Three Evaporation Estimation Methods in a Canadian Prairie Landscape Robert. *Hydrol. Process.* **2008**, *22*, 2801–2815. [[CrossRef](#)]
3. Fang, X.; Pomeroy, J.W.; Westbrook, C.J.; Guo, X.; Minke, A.G.; Brown, T. Prediction of Snowmelt Derived Streamflow in a Wetland Dominated Prairie Basin. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 1–16. [[CrossRef](#)]

4. Hayashi, M.; Van Der Kamp, G. Simple Equations to Represent the Volume-Area-Depth Relations of Shallow Wetlands in Small Topographic Depressions. *J. Hydrol.* **2000**, *237*, 74–85. [[CrossRef](#)]
5. Shook, K.; Pomeroy, J.; van der Kamp, G. The Transformation of Frequency Distributions of Winter Precipitation to Spring Streamflow Probabilities in Cold Regions; Case Studies from the Canadian Prairies. *J. Hydrol.* **2015**, *521*, 394–409. [[CrossRef](#)]
6. Eum, H.I.; Dibike, Y.; Prowse, T. Climate-Induced Alteration of Hydrologic Indicators in the Athabasca River Basin, Alberta, Canada. *J. Hydrol.* **2017**, *544*, 327–342. [[CrossRef](#)]
7. Hedstrom, N.R.; Granger, R.J.; Pomeroy, J.W.; Gray, D.M. Enhanced Indicators of Land Use Change and Climate Variability Impacts on Prairie Hydrology Using the Cold Regions Hydrological Model. In Proceedings of the 58th Eastern Snow Conference, Ottawa, ON, Canada, 17–19 May 2001; pp. 262–275.
8. Pattison-Williams, J.K.; Pomeroy, J.W.; Badiou, P.; Gabor, S. Analysis Wetlands, Flood Control and Ecosystem Services in the Smith Creek Drainage Basin: A Case Study in Saskatchewan, Canada. *Ecol. Econ. J.* **2018**, *147*, 36–47. [[CrossRef](#)]
9. Pomeroy, J.W.; Gray, D.M.; Brown, T.; Hedstrom, N.R.; Quinton, W.L.; Granger, R.J.; Carey, S.K. The Cold Regions Hydrological Model: A Platform for Basing Process Representation and Model Structure on Physical Evidence. *Hydrol. Process.* **2007**, *21*, 2650–2667. [[CrossRef](#)]
10. Mekonnen, M.A.; Wheeler, H.S.; Ireson, A.M.; Spence, C.; Davison, B.; Pietroniro, A. Towards an Improved Land Surface Scheme for Prairie Landscapes. *J. Hydrol.* **2014**, *511*, 105–116. [[CrossRef](#)]
11. Gray, D.M.; Landine, P.G. An Energy-Budget Snowmelt Model for the Canadian Prairies. *Can. J. Earth Sci.* **1988**, *25*, 1292–1303. [[CrossRef](#)]
12. Shook, K.; Pomeroy, J.W.; Spence, C.; Boychuk, L. Storage Dynamics Simulations in Prairie Wetland Hydrology Models: Evaluation and Parameterization. *Hydrol. Process.* **2013**, *27*, 1875–1889. [[CrossRef](#)]
13. Evenson, G.R.; Golden, H.E.; Lane, C.R.; D’Amico, E. An Improved Representation of Geographically Isolated Wetlands in a Watershed-Scale Hydrologic Model. *Hydrol. Process.* **2016**, *30*, 4168–4184. [[CrossRef](#)]
14. Brochero, D.; Anctil, F.; Gagné, C. Simplifying a Hydrological Ensemble Prediction System with a Backward Greedy Selection of Members—Part 2: Generalization in Time and Space. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 3327–3341. [[CrossRef](#)]
15. Ajami, N.K.; Duan, Q.; Gao, X.; Sorooshian, S. Multimodel Combination Techniques for Analysis of Hydrological Simulations: Application to Distributed Model Intercomparison Project Results. *J. Hydrometeorol.* **2006**, *7*, 755–768. [[CrossRef](#)]
16. Seiller, G.; Anctil, F.; Perrin, C. Multimodel Evaluation of Twenty Lumped Hydrological Models under Contrasted Climate Conditions. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 1171–1189. [[CrossRef](#)]
17. Velázquez, J.A.; Anctil, F.; Perrin, C. Performance and Reliability of Multimodel Hydrological Ensemble Simulations Based on Seventeen Lumped Models and a Thousand Catchments. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 2303–2317. [[CrossRef](#)]
18. Velázquez, J.A.; Anctil, F.; Ramos, M.H.; Perrin, C. Can a Multi-Model Approach Improve Hydrological Ensemble Forecasting? A Study on 29 French Catchments Using 16 Hydrological Model Structures. *Adv. Geosci.* **2011**, *29*, 33–42. [[CrossRef](#)]
19. Viney, N.R.; Bormann, H.; Breuer, L.; Bronstert, A.; Croke, B.F.W.; Frede, H.; Gräff, T.; Hubrechts, L.; Huisman, J.A.; Jakeman, A.J.; et al. Assessing the Impact of Land Use Change on Hydrology by Ensemble Modelling (LUCHEM) II: Ensemble Combinations and Predictions. *Adv. Water Resour.* **2009**, *32*, 147–158. [[CrossRef](#)]
20. Seiller, G.; Anctil, F.; Roy, R. Design and Experimentation of an Empirical Multistructure Framework for Accurate, Sharp and Reliable Hydrological Ensembles. *J. Hydrol.* **2017**, *552*, 313–340. [[CrossRef](#)]
21. Thibault, A.; Anctil, F.; Boucher, M.A. Accounting for Three Sources of Uncertainty in Ensemble Hydrological Forecasting. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 1809–1825. [[CrossRef](#)]
22. Antonetti, M.; Horat, C.; Sideris, I.V.; Zappa, M. Ensemble Flood Forecasting Considering Dominant Runoff Processes: I. Setup and Application to Nested Basins (Emme, Switzerland). *Nat. Hazards Earth Syst. Sci. Discuss.* **2018**, *5194*, 1–29. [[CrossRef](#)]
23. Alfieri, L.; Pappenberger, F.; Wetterhall, F.; Haiden, T.; Richardson, D.; Salamon, P. Evaluation of Ensemble Streamflow Predictions in Europe. *J. Hydrol.* **2014**, *517*, 913–922. [[CrossRef](#)]
24. Calvetti, L.; Pereira Filho, A.J. Ensemble Hydrometeorological Forecasts Using WRF Hourly QPF and Topmodel for a Middle Watershed. *Adv. Meteorol.* **2014**, *2014*. [[CrossRef](#)]



25. Fan, F.M.; Schwanenberg, D.; Kuwajima, J.; Assis, A. Ensemble Streamflow Predictions in the Três Marias Basin, Brazil. In Proceedings of the EGU General Assembly Conference, Vienna, Austria, 27 April–2 May 2014; Volume 16, p. 14191.
26. Liechti, K.; Zappa, M.; Fundel, F.; Germann, U. Probabilistic Evaluation of Ensemble Discharge Nowcasts in Two Nested Alpine Basins Prone to Flash Floods. *Hydrol. Process.* **2013**, *27*, 5–17. [\[CrossRef\]](#)
27. Pietroniro, A.; Fortin, V.; Kouwen, N.; Neal, C.; Turcotte, R.; Davison, B.; Verseghy, D.; Soulis, E.D.; Caldwell, R.; Evora, N.; et al. Development of the MESH Modelling System for Hydrological Ensemble Forecasting of the Laurentian Great Lakes at the Regional Scale. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1279–1294. [\[CrossRef\]](#)
28. Thiemi, V.; Pappenberger, F.; Thielen, J.; Gadain, H.; de Roo, A.; Bodis, K.; Del Medico, M.; Muthusi, F. Ensemble Flood Forecasting in Africa: A Feasibility Study in the Juba-Shabelle River Basin. *Atmos. Sci. Lett.* **2010**, *11*, 123–131. [\[CrossRef\]](#)
29. Zsótér, E.; Pappenberger, F.; Smith, P.; Emerton, R.E.; Dutra, E.; Wetterhall, F.; Richardson, D.; Bogner, K.; Balsamo, G. Building a Multimodel Flood Prediction System with the TIGGE Archive. *J. Hydrometeorol.* **2016**, *17*, 2923–2940. [\[CrossRef\]](#)
30. Fan, F.M.; Collischonn, W.; Meller, A.; Botelho, L.C.M. Ensemble Streamflow Forecasting Experiments in a Tropical Basin: The São Francisco River Case Study. *J. Hydrol.* **2014**, *519*, 2906–2919. [\[CrossRef\]](#)
31. Fan, F.M.; Schwanenberg, D.; Collischonn, W.; Weerts, A. Verification of Inflow into Hydropower Reservoirs Using Ensemble Forecasts of the TIGGE Database for Large Scale Basins in Brazil. *J. Hydrol. Reg. Stud.* **2015**, *4*, 196–227. [\[CrossRef\]](#)
32. Abaza, M.; Anctil, F.; Fortin, V.; Turcotte, R. A Comparison of the Canadian Global and Regional Meteorological Ensemble Prediction Systems for Short-Term Hydrological Forecasting. *Mon. Weather Rev.* **2013**, *141*, 3462–3476. [\[CrossRef\]](#)
33. Jasper, K.; Gurtz, J.; Lang, H. Advanced Flood Forecasting in Alpine Watersheds by Coupling Meteorological Observations and Forecasts with a Distributed Hydrological Model. *J. Hydrol.* **2002**, *267*, 40–52. [\[CrossRef\]](#)
34. Achleitner, S.; Schöber, J.; Rinderer, M.; Leonhardt, G.; Schöberl, F.; Kirnbauer, R.; Schönlaub, H. Analyzing the Operational Performance of the Hydrological Models in an Alpine Flood Forecasting System. *J. Hydrol.* **2012**, *412–413*, 90–100. [\[CrossRef\]](#)
35. De Roo, A.P.J.; Gouweleeuw, B.; Thielen, J.; Bartholmes, J.; Bongioannini-Cerlini, P.; Todini, E.; Bates, P.D.; Horritt, M.; Hunter, N.; Beven, K.; et al. Development of a European Flood Forecasting System. *Int. J. River Basin Manag.* **2003**, *1*, 49–59. [\[CrossRef\]](#)
36. Pappenberger, F.; Bartholmes, J.; Thielen, J.; Cloke, H.L.; Buizza, R.; de Roo, A. New Dimensions in Early Flood Warning across the Globe Using Grand-Ensemble Weather Predictions. *Geophys. Res. Lett.* **2008**, *35*, L10404. [\[CrossRef\]](#)
37. Demargne, J.; Wu, L.; Regonda, S.K.; Brown, J.D.; Lee, H.; He, M.; Seo, D.J.; Hartman, R.; Herr, H.D.; Fresch, M.; et al. The Science of NOAA's Operational Hydrologic Ensemble Forecast Service. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 79–98. [\[CrossRef\]](#)
38. Maxey, R.; Cranston, M.; Tavendale, A.; Buchanan, P. The Use of Deterministic and Probabilistic Forecasting in Countrywide Flood Guidance in Scotland. *Hydrol. Change World* **2012**, 1–7. [\[CrossRef\]](#)
39. Unduche, F.; Tolossa, H.; Senbeta, D.; Zhu, E. Evaluation of Four Hydrological Models for Operational Flood Forecasting in a Canadian Prairie Watershed. *Hydrol. Sci. J.* **2018**, *63*, 1–17. [\[CrossRef\]](#)
40. Hrachowitz, M.; Clark, M.P. HESS Opinions: The Complementary Merits of Competing Modelling Philosophies in Hydrology. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 3953–3973. [\[CrossRef\]](#)
41. Shrestha, R.R.; Dibike, Y.B.; Prowse, T.D. Modelling of Climate-Induced Hydrologic Changes in the Lake Winnipeg Watershed. *J. Great Lakes Res.* **2012**, *38* (Suppl. S3), 83–94. [\[CrossRef\]](#)
42. Hamill, T.M.; Bates, G.T.; Whitaker, J.S.; Murray, D.R.; Fiorino, M.; Galarneau, T.J.; Zhu, Y.; Lapenta, W. NOAA's Second-Generation Global Medium-Range Ensemble Reforecast Dataset. *Bull. Am. Meteorol. Soc.* **2013**, *94*, 1553–1565. [\[CrossRef\]](#)
43. Mahfouf, J.F.; Brasnett, B.; Gagnon, S. A Canadian Precipitation Analysis (CaPA) Project: Description and Preliminary Results. *Atmos. Ocean* **2007**, *45*, 1–17. [\[CrossRef\]](#)
44. Anderson, E. *Snow Accumulation and Ablation Model—SNOW-17*; National Weather Service: Silver Spring, MD, USA, 2006. [\[CrossRef\]](#)

45. Samuel, J.; Coulibaly, P.; Metcalfe, R.A. Estimation of Continuous Streamflow in Ontario Ungauged Basins: Comparison of Regionalization Methods. *J. Hydrol. Eng.* **2011**, *16*, 447–459. [[CrossRef](#)]
46. 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.* **1994**, *99*, 14415. [[CrossRef](#)]
47. Dibike, Y.; Eum, H.-I.; Prowse, T. Modelling the Athabasca Watershed Snow Response to a Changing Climate. *J. Hydrol. Reg. Stud.* **2018**, *15*, 134–148. [[CrossRef](#)]
48. Eum, H.I.; Dibike, Y.; Prowse, T. Uncertainty in Modelling the Hydrologic Responses of a Large Watershed: A Case Study of the Athabasca River Basin, Canada. *Hydrol. Process.* **2014**, *28*, 4272–4293. [[CrossRef](#)]
49. Eum, H.I.; Dibike, Y.; Prowse, T.; Bonsal, B. Inter-Comparison of High-Resolution Gridded Climate Data Sets and Their Implication on Hydrological Model Simulation over the Athabasca Watershed, Canada. *Hydrol. Process.* **2014**, *28*, 4250–4271. [[CrossRef](#)]
50. Kouwen, N. WATFLOOD: A Micro-Computer Based Flood Forecasting System Based on Real-Time Weather Radar. *Can. Water Resour. J.* **1988**, *13*, 62–77. [[CrossRef](#)]
51. Haghnegahdar, A.; Tolson, B.A.; Davison, B.; Seglenieks, F.R.; Klyszejko, E.; Soulis, E.D.; Fortin, V.; Matott, L.S. Calibrating Environment Canada's MESH Modelling System over the Great Lakes Basin. *Atmos. Ocean* **2014**, *52*, 281–293. [[CrossRef](#)]
52. Newman, A.J.; Mizukami, N.; Clark, M.P.; Wood, A.W.; Nijssen, B.; Nearing, G.; Newman, A.J.; Mizukami, N.; Clark, M.P.; Wood, A.W.; et al. Benchmarking of a Physically Based Hydrologic Model. *J. Hydrometeorol.* **2017**, *18*, 2215–2225. [[CrossRef](#)]
53. Hopkinson, R.F.; Mckenney, D.W.; Milewska, E.J.; Hutchinson, M.F.; Papadopol, P.; Vincent, A.L.A. Impact of Aligning Climatological Day on Gridding Daily Maximum-Minimum Temperature and Precipitation over Canada. *J. Appl. Meteorol. Climatol.* **2011**, *50*, 1654–1665. [[CrossRef](#)]
54. Hutchinson, M.F.; McKenney, D.W.; Lawrence, K.; Pedlar, J.H.; Hopkinson, R.F.; Milewska, E.; Papadopol, P. Development and Testing of Canada-Wide Interpolated Spatial Models of Daily Minimum-Maximum Temperature and Precipitation for 1961–2003. *J. Appl. Meteorol. Climatol.* **2009**, *48*, 725–741. [[CrossRef](#)]
55. Côté, J.; Desmarais, J.G.; Gravel, S.; Méthot, A.; Patoine, A.; Roch, M.; Staniforth, A. The Operational CMC-MRB Global Environmental Multiscale (GEM) Model. Part II: Results. *Mon. Weather Rev.* **1998**, *126*, 1397–1418. [[CrossRef](#)]
56. Friedl, M.; Sulla-Menashe, D. MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500 m SIN Grid V006 [Data Set]; NASA EOSDIS Land Processes DAAC: Washington, DC, USA, 2015. [[CrossRef](#)]
57. FAO; JRC; IIASA; ISRIC; ISS-CAS. *Harmonized World Soil Database—Version 1.1*; FAO: Rome, Italy; IIASA: Laxenburg, Austria, 2009.
58. Hamman, J.; Nijssen, B.; Clark, E.; Matthews, D.; Veerman, B. *UW-Hydro/RVIC: RVIC 1.1.1*, v1.1.1. 2017. [[CrossRef](#)]
59. Lohmann, D.; Nolte-Holube, R.; Raschke, E. A Large-Scale Horizontal Routing Model to Be Coupled to Land Surface Parametrization Schemes. *Tellus Ser. A Dyn. Meteorol. Oceanogr.* **1996**, *48*, 708–721. [[CrossRef](#)]
60. Tolson, B.A.; Shoemaker, C.A. Dynamically Dimensioned Search Algorithm for Computationally Efficient Watershed Model Calibration. *Water Resour. Res.* **2007**, *43*, 1–16. [[CrossRef](#)]
61. Ilampooranan, I.; Van Meter, K.J.; Basu, N.B. A Race against Time: Modeling Time Lags in Watershed Response. *Water Resour. Res.* **2019**, *55*, 3941–3959. [[CrossRef](#)]
62. Aliyari, F.; Bailey, R.T.; Tasdighi, A.; Dozier, A.; Arabi, M.; Zeiler, K. Coupled SWAT-MODFLOW Model for Large-Scale Mixed Agro-Urban River Basins. *Environ. Model. Softw.* **2019**, *115*, 200–210. [[CrossRef](#)]
63. Rokaya, P.; Wheeler, H.; Lindenschmidt, K.E. Promoting Sustainable Ice-Jam Flood Management along the Peace River and Peace-Athabasca Delta. *J. Water Resour. Plan. Manag.* **2019**, *145*, 1–12. [[CrossRef](#)]
64. Krogh, S.A.; Pomeroy, J.W. Impact of Future Climate and Vegetation on the Hydrology of an Arctic Headwater Basin at the Tundra-Taiga Transition. *J. Hydrometeorol.* **2019**, *20*, 197–215. [[CrossRef](#)]
65. Sharma, S.; Siddique, R.; Reed, S.; Ahnert, P.; Mejia, A. Hydrological Model Diversity Enhances Streamflow Forecast Skill at Short- to Medium-Range Timescales. *Water Resour. Res.* **2019**, *55*, 1510–1530. [[CrossRef](#)]
66. Lahmers, T.M.; Gupta, H.; Castro, C.L.; Gochis, D.J.; Yates, D.; Dugger, A.; Goodrich, D.; Hazenberg, P. The structure of the W. hydrologic model for semiarid environments. Enhancing the Structure of the WRF-Hydro Hydrologic Model for Semiarid Environments. *J. Hydrometeorol.* **2019**, *20*, 691–714. [[CrossRef](#)]

67. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the Mean Squared Error and NSE Performance Criteria: Implications for Improving Hydrological Modelling. *J. Hydrol.* **2009**, *377*, 80–91. [\[CrossRef\]](#)
68. Coulibaly, P.; Anctil, F.; Bobée, B. Multivariate Reservoir Inflow Forecasting Using Temporal Neural Networks. *J. Hydrol. Eng.* **2001**, *6*, 367–376. [\[CrossRef\]](#)
69. Amengual, A.; Homar, V.; Romero, R.; Alonso, S.; Ramis, C.; Amengual, A.; Homar, V.; Romero, R.; Alonso, S.; Ramis, C. A Statistical Adjustment of Regional Climate Model Outputs to Local Scales: Application to Platja de Palma, Spain. *J. Clim.* **2012**, *25*, 939–957. [\[CrossRef\]](#)
70. Wilks, D. *Statistical Methods in the Atmospheric Sciences*, 2nd ed.; Academic Press: Cambridge, MA, USA, 2006; Volume 91.
71. Hersbach, H. Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems. *Weather Forecast.* **2000**, *15*, 559–570. [\[CrossRef\]](#)
72. Bradley, A.A.; Schwartz, S.S. Summary Verification Measures and Their Interpretation for Ensemble Forecasts. *Mon. Weather Rev.* **2011**, *139*, 3075–3089. [\[CrossRef\]](#)
73. Muhammad, A.; Stadnyk, T.; Unduche, F.; Coulibaly, P.; Muhammad, A.; Stadnyk, T.A.; Unduche, F.; Coulibaly, P. Multi-Model Approaches for Improving Seasonal Ensemble Streamflow Prediction Scheme with Various Statistical Post-Processing Techniques in the Canadian Prairie Region. *Water* **2018**, *10*, 1604. [\[CrossRef\]](#)
74. Pappenberger, F.; Ramos, M.H.; Cloke, H.L.; Wetterhall, F.; Alfieri, L.; Bogner, K.; Mueller, A.; Salamon, P. How Do I Know If My Forecasts Are Better? Using Benchmarks in Hydrological Ensemble Prediction. *J. Hydrol.* **2015**, *522*, 697–713. [\[CrossRef\]](#)
75. Hsu, W.-R.; Murphy, A.H. The Attributes Diagram A Geometrical Framework for Assessing the Quality of Probability Forecasts. *Int. J. Forecast.* **1986**, *2*, 285–293. [\[CrossRef\]](#)
76. Atger, F. The Skill of Ensemble Prediction Systems. *Mon. Weather Rev.* **1999**, *127*, 1941–1953. [\[CrossRef\]](#)
77. Mason, S.J.; Graham, N.E. Areas beneath the Relative Operating Characteristics (ROC) and Relative Operating Levels (ROL) Curves: Statistical Significance and Interpretation. *Q. J. R. Meteorol. Soc.* **2002**, *128*, 2145–2166. [\[CrossRef\]](#)
78. Brown, J.D.; Demargne, J.; Seo, D.J.; Liu, Y. The Ensemble Verification System (EVS): A Software Tool for Verifying Ensemble Forecasts of Hydrometeorological and Hydrologic Variables at Discrete Locations. *Environ. Model. Softw.* **2010**, *25*, 854–872. [\[CrossRef\]](#)
79. Chang, H.-L.; Yang, S.-C.; Yuan, H.; Lin, P.-L.; Liou, Y.-C.; Chang, H.-L.; Yang, S.-C.; Yuan, H.; Lin, P.-L.; Liou, Y.-C. Analysis of the Relative Operating Characteristic and Economic Value Using the LAPS Ensemble Prediction System in Taiwan. *Mon. Weather Rev.* **2015**, *143*, 1833–1848. [\[CrossRef\]](#)
80. Mason, S.J.; Graham, N.E. Conditional Probabilities, Relative Operating Characteristics, and Relative Operating Levels. *Weather Forecast.* **1999**, *14*, 713–725. [\[CrossRef\]](#)
81. Yapo, P.O.; Gupta, H.V.; Sorooshian, S. Automatic Calibration of Conceptual Rainfall-Runoff Models: Sensitivity to Calibration Data. *J. Hydrol.* **1996**, *181*, 23–48. [\[CrossRef\]](#)
82. Reed, S.; Koren, V.; Smith, M.; Zhang, Z.; Morea, F.; Seo, D.J. Overall Distributed Model Intercomparison Project Results. *J. Hydrol.* **2004**, *298*, 27–60. [\[CrossRef\]](#)
83. Duan, Q.; Schaake, J.; Andréassian, V.; Franks, S.; Goteti, G.; Gupta, H.V.; Gusev, Y.M.; Habets, F.; Hall, A.; Hay, L.; et al. Model Parameter Estimation Experiment (MOPEX): An Overview of Science Strategy and Major Results from the Second and Third Workshops. *J. Hydrol.* **2006**, *320*, 3–17. [\[CrossRef\]](#)
84. Nasonova, O.N.; Gusev, Y.M. Can a Land Surface Model Simulate Runoff with the Same Accuracy as a Hydrological Model? *IAHS AISH Public.* **2007**, *313*, 258–265.
85. Maurer, E.P.; Brekke, L.D.; Pruitt, T. Contrasting Lumped and Distributed Hydrology Models for Estimating Climate Change Impacts on California Watersheds. *J. Am. Water Resour. Assoc.* **2010**, *46*, 1024–1035. [\[CrossRef\]](#)
86. Smith, M.B.; Seo, D.J.; Koren, V.I.; Reed, S.M.; Zhang, Z.; Duan, Q.; Morea, F.; Cong, S. The Distributed Model Intercomparison Project (DMIP): Motivation and Experiment Design. *J. Hydrol.* **2004**, *298*, 4–26. [\[CrossRef\]](#)
87. Koren, V.; Reed, S.; Smith, M.; Zhang, Z.; Seo, D.-J. Hydrology Laboratory Research Modeling System (HL-RMS) of the US National Weather Service. *J. Hydrol.* **2004**, *291*, 297–318. [\[CrossRef\]](#)

88. National Weather Service. *Conceptualization of the Sacramento Soil Moisture Accounting Model Introduction*; NWS: Silver Spring, MD, USA, 2002.
89. Bergström, S. *Development of a Conceptual Deterministic Rainfall—Runoff Model*; Nordic Hydrology: Gothenburg, Stockholm, 1978.
90. Leach, J.M.; Kornelsen, K.C.; Coulibaly, P. Assimilation of Near-Real Time Data Products into Models of an Urban Basin. *J. Hydrol.* **2018**, *563*, 51–64. [[CrossRef](#)]
91. Razavi, T.; Coulibaly, P. Improving Streamflow Estimation in Ungauged Basins Using a Multi-Modelling Approach. *Hydrol. Sci.* **2016**, *61*, 2668–2679. [[CrossRef](#)]
92. Razavi, T.; Coulibaly, P. An Evaluation of Regionalization and Watershed Classification Schemes for Continuous Daily Streamflow Prediction in Ungauged Watersheds. *Can. Water Resour. J. Rev. Can. Ressour. Hydr.* **2017**, *42*, 2–20. [[CrossRef](#)]
93. Han, S.; Coulibaly, P. Probabilistic Flood Forecasting Using Hydrologic Uncertainty Processor with Ensemble Weather Forecasts. *J. Hydrometeorol.* **2019**, *20*, 1379–1398. [[CrossRef](#)]
94. Han, S.; Coulibaly, P.; Biondi, D. Assessing Hydrologic Uncertainty Processor Performance for Flood Forecasting in a Semiurban Watershed. *J. Hydrol. Eng.* **2019**, *24*, 05019025. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).