CGU HS Committee on River Ice Processes and the Environment
22nd Workshop on the Hydraulics of Ice-Covered Rivers
Canmore, Alberta, Canada, July 9-12, 2023.

Comparison of seasonal and year-long calibration approaches for hydrological modelling of winter discharge on the Chaudière River, Quebec

Kh Rahat Usman[1], Tadros Ghobrial[1], and François Anctil[1]

[1]Civil and Water Engineering Department, Laval University, Québec City, G1V 0A6 Canada
kh-rahat.usman.1@ulaval.ca;
tadros.ghobrial@gci.ulaval.ca;
francois.Anctil@gci.ulaval.ca

Discharge estimation and forecasting in ice covered rivers is challenging due to the continuous spatio-temporal variation in river ice conditions. Hydrological modelling is a common tool for discharge forecasting; however, during ice cover season, this approach may suffer from various sources of uncertainty arising from meteorological forcing, hydrological model structure, model states and uncertainty in calibration data. The current study takes a probabilistic approach for estimating and forecasting winter discharge to account for the different sources of uncertainty in hydrological modelling through a multi-modelling framework HOOPLA, ensemble meteorological forecasts and data assimilation. It also evaluates the potential of different calibration approaches such as season-based calibration against year-long calibration. This work is part of developing a broader framework for under ice discharge estimation and forecasting through coupled hydrological and hydraulic modelling for the Chaudière River in Quebec. The HOOPLA framework was calibrated and validated on 10-year observational datasets from 2008 to 2018. The performance of the calibrated and validated framework was then assessed by hindcasting the 2018-2019 winter discharge data. The performance matrices used for forecast evaluation include Mean Continuous Ranked Probability Score (MCRPS), Root Mean Square Error (RMSE) and Relative Bias. Our work shows that seasonal hydrological models (considering winter only data) perform relatively better and yield a lower MCRPS. However, they are marred by over forecasting and larger ensemble spread compared to the conventional hydrological modelling approach indicating higher degree of uncertainty in the streamflow forecasts.

1 Introduction

Agencies across Canada, responsible for collecting and disseminating hydrometric data face an uphill task when dealing with winter discharge estimates under ice cover conditions. As the ice cover forms over a stream, an additional roughness layer is exerted on the flow by the floating ice cover, which changes with time and along the reach. Stage-Discharge relationships or rating curves are commonly developed for determining discharge under open water conditions however, the presence of an ice cover changes the flow resistance, and consequently the river stage hence, rendering the open water rating curves useless (Teal, Ettema et al. 1996, Fulton, Henneberg et al. 2018). Direct measurement is the best-known method till date for measuring under ice discharge however, this technique is not only costly and resources intensive but also involves risks for the hydrographers. These difficulties have led to the lack of real-time information as well as missing or highly uncertain winter hydrometric records.

Hydrological modelling approaches have been tested for the simulation of winter discharge (Hamilton, Hutchinson et al. 2000, Turcotte, Favre et al. 2005, Levesque, Anctil et al. 2008). Hamilton, Hutchinson et al. (2000) found that hydrological modelling produced reasonable estimates for winter discharge however, failed to capture discharge variability over the season. Turcotte, Favre et al. (2005) noted that there is potential in seeking improvements in hydrological model calibration by incorporating winter streamflow gauging data to improve their performance for winter discharge estimates. Levesque, Anctil et al. (2008) studied the hydrological response of two small catchments under snowmelt and rainfall conditions. They applied different calibration methods for year-long streamflow simulations and identified the significance of seasonal effects on simulations. They noted that calibrating their SWAT model for only winter data improved winter simulations. Singh and Bárdossy (2012) found that calibrating hydrological models over a small number of carefully selected events can be sufficient and produce satisfactory results. Furthermore, in a review of methodologies for assessing flood frequencies in ice covered rivers, Beltaos (2021) also suggested testing *"targeted model calibrations, i.e., ascertaining good model performance during a key period"* for taking up hydrological modelling approach under river ice conditions.

Mostly these studies simulated winter discharge in a deterministic manner. The results from these studies warrant further investigation into hydrological modelling approach aimed at improving the modelling process as well as accounting for the uncertainties in the modelling chain. Hydrological simulation and forecasting attracts uncertainties from four main sources namely: (i) model structure and parameters; (ii) observations; (iii) meteorological forcings; and (iv) initial conditions (Thiboult, Anctil et al. 2016). A popular approach to account for these uncertainties is the ensemble based methods (Boelee, Lumbroso et al. 2019) that rely upon simulating several possible outcomes for the modelled/forecasted variable by combing different model structures, model states scenarios and forcing scenarios. Thereby, generating useful information for complete description of uncertainty associated with the modelling process.

This study attempts to build on some of the findings of previous research addressing winter discharge estimation and forecasting. The main objective of this study is to evaluate the performance of different model calibration approaches for winter discharge estimation and forecasting in a probabilistic manner. To achieve this objective, three model calibration systems have been tested: (i) Winter system comprising of calibrated parameters obtained by calibrating only on ice affected data; (ii) Year-long system consisting of model calibration achieved by utilizing the year-long observation time series; and (iii) Open water system where the model was calibrated based on segment of discharge time-series unaffected by river ice conditions. An ensemble-based approach has been adopted to account for the different sources of uncertainty in the modelling chain. This consists of a multi-modelling framework (HOOPLA) to address the issue of model structure uncertainty; ensemble meteorological forcings accounting for the forcing data uncertainty; and the application of data assimilation catering the uncertainty associated with observations and model initial conditions for each forecast.

2 Methodology

2.1 Study Area

The study is conducted on the upstream sub-catchment of the Chaudière River basin (referred to as the Upper Chaudière) located in the South-East of Québec, Canada (Figure 1). The Chaudière River originates from Mégantic Lake and drains into the Saint-Lawrence River south of Quebec City, covering a 185 km long reach and draining a total watershed of 6694 km^2. As shown in Figure 1, an ice control structure, called Sartigan Dam, is built on the downstream end of the Upper Chaudière River. The Sartigan Dam is a run-of-the-river dam, i.e., built primarily to intercept ice runs and doesn't affect the natural regime of the river discharge. The Upper Chaudière drains a total area of 3074 $km^2$and the Centre d'expertise hydrique du Québec (CEHQ) has a hydrometric station (Station ID: 023429) at its mouth, just downstream of Sartigan Dam. The climate of the catchment is classified as humid continental (Dfb) according to the Köppen classification (Kottek, Grieser et al. 2006), a dominant climate regime for Southern Québec. The average monthly temperatures show a high degree of seasonal variation with below freezing temperatures (average -6 °C and min -12.6 °C) between November-March, and moderate temperatures (average 18 °C, high 25 °C) during the summers. The average annual precipitation is estimated to be 1031.5 mm shared between rain (824.9 mm) and snow (202.6 mm) (Ministère de l'Environnement 2023). Monthly precipitation averages do not exhibit significant variations, however between June and August slight intensity in precipitation exists. The hydrological regime for the catchment can be classified as nivo-pluvial since the spring flood caused by snowmelt and precipitation is dominant, followed by flooding in autumn season due to excess rainfall (Ricard, Lucas-Picher et al. 2022).

The Upper Chaudière River has a steep profile with a slope of 2.5 m/km while in the Intermediate Chaudière, downstream of the study catchment, the profile becomes gentle with a slope of 0.5 m/km (Bessar 2021). Ice jam floods are common in the Intermediate Chaudière downstream of Sartigan Dam where the river profile is flatter, and a number of structures (culverts and bridges) have been constructed.

2.2 Data

The data used for this study consists of hydro-meteorological observations and meteorological forecasts. The meteorological observations consist of precipitation and temperature (Min, Max, and Mean) while hydrometric observations consist of river discharge. The time-series for all the parameters extends from January 2007 to December 2019 at a 3h timestep. Centre d'expertise hydrique du Québec (CEHQ) provided gridded meteorological data at 0.1^{o} – resolution grid, constructed through ordinary kriging from point observations obtained from a dense network of climatic stations across Québec (Valdez, Anctil et al. 2022). This observed meteorological data was used for model calibration and validation.

River discharge data is obtained from CEHQ station 023429. During open water conditions the data available from CEHQ is at a 15-minute temporal resolution, however during winters when the river undergoes freeze up, typically from December until March, CEHQ produces a corrected average daily discharge time-series for this period (corrected by applying a backwater correction factor to the open water rating curves). This introduces a source of uncertainty in the observational data. Moreover, missing data in observational time-series is also not a unique problem in hydrology. It is worth noting here that within the available hydrometric observation time-series no observations were available from 27-12-2015 to 14-03-2016, which constitutes nearly an entire ice affected discharge season. The time step for hydrological modelling was set to be 3-hours, the average daily discharges for the winter period were down scaled to a 3-hour resolution through linear interpolation. For the remaining period in each year the 15-minute hydrometric data was converted to 3-hour resolution by simply averaging the observations over the timestep.

The meteorological forecasts, used for forecasting runoff, in this study were produced by the European Center for Medium range Weather Forecasts (ECMWF) and were retrieved from the TIGGE database. The forecast data used in this study is 94 days long starting from January 1st to April 4th, 2019. The forecasts data consists of precipitation and temperatures. The forcing ensemble has 50 exchangeable members (i.e., a set of 50 forecasts issued for a each timestep over the forecast horizon and each forecast having equal likelihood of occurrence) at a 6-hours timestep, and a 10-days forecast horizon. For simplicity the forecast horizon considered in this study is 6.5 days. The ECMWF forecasts have a 0.5^{o} spatial resolution. To ensure a finer description of the catchment meteorology the raw forecasts were downscaled to a spatial resolution of 0.1^{o} through bilinear interpolation (Gaborit, Anctil et al. 2013). The meteorological forecasts for each ensemble member were averaged over the catchment to obtain a representative forecast from the downscaled grid points (Thiboult, Anctil et al. 2016).

2.3 HOOPLA

The HydrOlOgical Prediction Laboratory (HOOPLA) is a multi-model hydrological modeling framework that consists of 20 conceptual lumped hydrological models (Thiboult 2019). The initial selection of models was done by Perrin (2000) based on their conceptual and structural diversity and later revised by Seiller, Anctil et al. (2012). The models included in HOOPLA vary from low to moderate complexity with 4 to 10 calibration parameters and 2 to 7 reservoirs for hydrological process description (Thiboult, Anctil et al. 2016). This structural diversity among the models is a key feature of the framework that provides an ensemble like description for storage and routing processes. The models implemented in HOOPLA are either in their original form or slightly modified to fit them within the common framework (i.e., to meet the input and computation requirements of the common framework). The hydrological processes are calculated at a catchment scale and the parameters are uniform over the entire catchment. Thiboult and Anctil (2015) have demonstrated that the framework’s performance is satisfactory and competitive against more complex models.

The HOOPLA framework is provided with a snow accounting routine: Cemaneige (Valéry, Andréassian et al. 2014) and a module to calculate potential evaportranspiration (PET). The provision of these external modules led to the omission of innate snow accounting and PET routines of individual models integrated in the framework, thereby maintaining a common representation of the hydrological processes. The snow accounting routine Cemaneige is a simple degree-day snow accounting routine with a spatial discretization of the catchment. It divides the catchment into five equal-area elevation bands to compute accumulation and melt within each elevation band. The two calibration parameters i.e., snowmelt factor and cold-content factor, are calibrated individually for each model using an objective function based on observed discharge (Thiboult 2019).

Furthermore, the HOOPLA framework is equipped with Data Assimilation (DA) capabilities which make it relevant for streamflow forecasting applications. Data assimilation, nowadays, is considered as a prerequisite for streamflow forecasting due to its proven capabilities of enhancing the streamflow simulation quality (Liu, Weerts et al. 2012, Boucher, Quilty et al. 2020). Two probabilistic, sequential DA schemes: (i) Ensemble Kalman Filter (EnKF); (ii) Sequential Importance Resampling filter (SIR) belonging to the particle filters (PF) have been integrated into the HOOPLA framework. The primary objective of DA in hydrological simulations and forecasting is to incorporate the new information i.e., the observations, as they become available. DA compares the observations with the model predictions and the difference is used to reinitialize the model with the aim to reduce the difference by providing improved initial conditions for the next timestep. The Ensemble Kalman Filter (EnKF) is a sequential assimilation method, which strives to find the set of states that statistically best fit the observations. The EnKF takes a Monte Carlo approach to identify error covariances required to compute the updated states (Thiboult 2015).

Calibration within such a framework can be a daunting task since models vary in structure and number of calibration parameters. This challenge is tackled in HOOPLA by the provision of two automatic calibration algorithms: (i) the Shuffled Complex Evolution (SCE) (Duan, Sorooshian et al. 1992), and (ii) the Dynamically Dimensioned Search (DSS) (Tolson and Shoemaker 2007) algorithm. The algorithms are iterative and global i.e., seek optimal parameter set within the parameter space. For more details on HOOPLA the readers are encouraged to consult the User Manual: HOOPLA version 1.0.2 (Thiboult 2019).

2.4 Model Calibration and Validation

Three calibration systems were developed for forecasting streamflow in winter months under river ice conditions. The available data was categorized into three sets: (i) Ice affected discharge termed as Winter system; (ii) Full set of continuous hydrometric data containing both open water and ice affected discharge data as year-long system; and (iii) Open water discharge time-series i.e., discharge time-series for periods with little to no influence of river ice, termed as the Open water system. The time-series used in this study was divided into two equal segments as proposed by KlemeŠ (1986). The first segment consisted of records from 2008 to 2012 and the second segment consisted of records from 2014 to 2018. The starting and end years are inclusive making a 5-year long time-series for each segment. The years 2007 and 2013 were used for model spin up for calibration and validation respectively.

All 20 models included in HOOPLA framework were used in this study. Automatic calibration using the SCE algorithm was performed and the modified Kling-Gupta efficiency (KGEm) was computed (Kling, Fuchs et al. 2012) as the calibration objective function. The same metric was used for evaluating model performance during the validation period.

2.5 Streamflow forecasting

Streamflow forecasting was done for the winter of 2019. The period considered ranged from January 01 to April 04, 2019. The forecasting approach made use of DA implemented through the EnKF algorithm. EnKF requires a prior knowledge of uncertainty associated with the observations to estimate the probability distribution of the model state variables. EnKF's performance is highly dependent on its hyperparameters representing uncertainty associated with model inputs and outputs as well as the selected state variables upon which it is applied. For this study, the settings derived from Thiboult, Anctil et al. (2016) were used. Thiboult and Anctil (2015) performed extensive testing to determine the optimal set of EnKF parameters over a region that includes the study catchment. The EnKF settings used assign a 50% standard deviation of the mean with gamma law to precipitation, a 10% standard deviation with normal distribution to the streamflows and 2 ^{o}C standard deviation to temperature. The DA ensemble size was kept to 50 members only. The overall forecast ensemble takes a size of 50,000 (20x50x50) members with a forecast horizon of 6.5 days (156 hours).

Forecast evaluation was done considering three main metrices namely: Continuous Ranked Probability Score (CRPS); Root Mean Square Error (RMSE); and Bias. Forecast reliability is considered by comparing RMSE and ensemble spread (Fortin, Abaza et al. 2014).

3 Results and Discussion

3.1 Calibration and validation

Figure 2 shows the results of the HOOPLA framework's calibration and validation. The models in HOOPLA reproduce the observed hydrograph very well. They capture most of the peak flows and both in terms of magnitude and timing. Figure 3 shows the results for the KGEm performance metric of the three systems during the calibration and validation (for year-long hydrograph) for all 20 models. Table 1 summarizes the average, minimum and maximum values of the objective function for each system. We calculated the objective function for both year-long hydrograph and winter specific hydrograph. In general, all three systems produce a very good agreement during the calibration and validation periods. The results summarized in the Figure 3 show that the average KGEm value for both calibration and validation periods is high and comparable indicating the appropriateness of the modelling framework for application in the given basin. It is also interesting to see that during the validation for year-long hydrograph, on average the models either perform equally or slightly better than the calibration period. The evaluation of precipitation records showed no marked difference between the calibration and validation period. The average annual precipitation during the two periods differed only by 38 mm i.e., 1188.17 mm average annual precipitation during the calibration period while 1150.13 mm during the validation period. The only identifiable difference in precipitation during the two periods was in precipitation pattern. Calibration period had a couple of years with precipitation approximately 13% higher than the average whereas for the validation period the wettest year received only 4% additional precipitation. This consistency in precipitation periods explains why some models show better performance during the validation period.

In case of validation against winter specific hydrograph, KGEm evaluations show that targeted calibration system i.e., the winter system produces better results, with an average KGEm of 0.88, than year-long or open water systems, whose averages are 0.85 and 0.79 respectively. In this case the open water system performs worse than the other two. This finding is similar to the findings of Levesque, Anctil et al. (2008). However, contrary to the notion that targeted calibration may worsen year-long model performance this system performed comparable to the other two systems. The average KGEm for year-long simulation of the winter system was found to be 0.92, which is slightly better than the open water system ($KGEm_{av}$ 0.91).

The year-long system also showed good winter performance with an average $KGEm_{av}$ of 0.85, it's year-long performance met the expectations that this could be the best performing system. This system showed high accuracy with a $KGEm_{av}$ of 0.94. The open water system also showed good performance for the year-long simulations however, its winter specific application is inferior to the other two systems.

3.2 Streamflow forecasting

Figure 4 shows an example of a 6.5 days forecast hydrograph starting from 2019-01-01 at 09:00 till 2019-01-07 at 18:00. The ensemble mean, 1st and 3rd quartiles also known as the interquartile range (25th and 75th percentiles of the ensemble), and the lower and upper limits (10th and 90th percentiles) of the ensemble have been plotted for the three systems. The ensemble mean produced by the winter system (Figure 4a) follows closely the observed hydrograph for the first few days before starting to overestimate the discharge on the 5th day. The year-long system (Figure 4b) also indicates a smaller error between the observation and forecast as compared to the open water system (Figure 4c). However, all three systems indicate discharge increasing progressively whereas the observed hydrograph is a receding one as the forecast horizon is approached. The upper and lower limits of the ensemble show that prediction uncertainty increases significantly over the forecast horizon in all three systems. The interquartile range (IQR) of the predictions, however, remains consistent over the forecast horizon. For all three systems, the ensemble mean lies within the IQR for a period of 24 hours (8 time steps) and afterwards extends above the IQR indicating right skew (also known as positive skew) and outliers in the ensemble predictions. This was confirmed through a histogram analysis of the ensemble distribution (not presented in this paper). The presence of extreme values tends to pull the ensemble mean outside the IQR as we approach the forecast horizon and thereby reducing the confidence in the predictions at higher lead times.

3.3 CRPS

CRPS is one of the most common metrics used for forecast evaluation. It is a measure of overall accuracy and reliability of the forecast. Mean CRPS (MCRPS) is simply the average of CRPS for each lead time over the entire evaluation period. CRPS calculates the average distance between the cumulative distribution function (CDF) of the forecast probabilities and the CDF of observations (calculated by Heaviside step function). CRPS has no upper limit while on the lower limit a CRPS value of 0 represents a perfect forecast (Valdez, Anctil et al. 2022). Smaller values of MCRPS indicate a close match between observations and forecasts. Figure 5 presents the MCRPS calculated at each lead time over the evaluation period for the three systems. The analysis shows that all three systems exhibit relatively similar performance. On a relative scale, the winter system performs the best, having the lowest MCRPS for all lead times, with MCRPS between 14.43 to 18.86 m^3/s, followed by the year-long system where the MCRPS ranges from 15.16 to 20.18 m^3/s. The MCRPS for the open water system ranges between 16 to 21.72 m^3/s. The MCRPS values remain in a tight range of 4 to 6 m^3/s for all the systems during the evaluation period indicating consistency in the predictions of the system. It also highlights the fact that discharge variability under river ice conditions is relatively low.

The shape of the MCRPS lines indicate the presence of a diurnal cycle in the hydrological forecasts. The point of inflection corresponds to 00:00 EST. This diurnal cycle is more pronounced in the RMSE, Spread and BIAS analysis (Figures 6 to 9). The presence of diurnal cycle in the forecasts indicates streamflow variability, however, this is difficult to confirm since observations are available on a daily resolution and mere statistical downscaling might not be able to capture the actual streamflow variation during the day.

3.4 RMSE and Spread

The Root Mean Square Error (RMSE) is another common metric used for gauging the performance of a predictive system. In ensemble forecasting systems, RMSE is computed as the difference between the ensemble members and the observations. It represents the average error between predictions and observations. We computed average RMSE against each lead time for the evaluation period as shown in Figure 6. The performance of all three systems is quite similar. On a relative scale, the error analysis shows that the winter system has slightly higher RMSE when compared to the other two systems. The year-long and open water systems perform almost equally. The RMSE for the three systems ranges between 55 to 120 m^3/s. This is significantly high error when compared to the order of observed discharge during the evaluation period. The average discharge in this period is 42.95 m^3/s.

The error keeps on increasing till the end of the forecast horizon which is expected since the uncertainty in meteorological forecast increases with each lead time. RMSE exhibits the diurnal cycle over the forecast horizon similar to the one seen in MCRPS plots, however, here it is more pronounced. An inspection of the meteorological forecasts revealed that precipitation and temperature forecasts were pulsating. Additionally, the lumped nature of the rainfall-runoff models also plays a role in causing this pulsating behavior whereas in nature the hydrological processes are spatially discretized.

Ensemble spread is calculated following the procedure proposed by Fortin, Abaza et al. (2014). The ensemble spread represents dispersion among the ensemble members. The calculated ensemble spread is presented in Figure 7. The ensemble spread closely follows the RMSE plot. The comparison of the two figures shows that the ensemble spread and RMSE are of the same order. The performance order of the three systems in ensemble spread calculations is identical to the RMSE analysis i.e., the year-long and open water systems having almost equal spread at each lead time while the winter system has slightly larger spread.).

3.5 Reliability

One core aspect of ensemble forecasting is to establish its reliability. Reliability refers to confidence in the performance of an ensemble forecasting system i.e., the ability of the ensemble predictions to accurately produce observable events. We employ a simple, yet popular graphical method to test the reliability of modelling systems. This approach consists of comparing the ensemble mean RMSE to the ensemble spread.

This analysis is presented in Figure 8 and shows that the forecast reliability of all three systems is very good. Spread and RMSE are of the same order and follow the same trend over the forecast horizon. This indicates that the three systems are able to describe the forecast uncertainty quite well and are able to capture variability in the discharge. The spread is slightly below the RMSE but not as significantly as to conclude under-dispersion. This also gives insights for logically reducing the ensemble size to further improve the reliability of the system. This could be done by carefully selecting a limited number of models from HOOPLA that best describe the catchment's hydrology.

3.6 BIAS

As a final test we evaluated the relative bias of the three systems. The relative bias is the ratio between the mean forecast streamflow to the observations at each lead time. A value of 1 would indicate low bias and values greater than 1 show over-forecasting and lower than 1 indicate under-forecasting. Figure 9 summarizes the relative bias evaluation of the three systems. The open water system yields the lowest relative bias as compared to the other two systems. Its relative bias is in the range of 1.2 to 1.83. For the year-long system the relative bias ranges from 1.3 to 1.97 while for the winter system the relative bias ranges from 1.5 to 2.2. This analysis shows that the winter system over-forecast the discharge when compared to the other two systems. The other two systems are nearly comparable to each other. The high relative bias in the winter system hints that it could be suffering from model overfitting problems. Also, uncertainty in the observations make it difficult to definitively address the over-forecasting issue. A simple approach to solve this issue could be to establish a bias correction factor.

4 Conclusion

In this study we considered three different approaches for calibration of a multi-model hydrologic modelling framework (HOOPLA) to improve the estimation and forecasting of winter streamflow under river ice conditions. The calibration and validation process showed good agreement with observations. Targeted winter calibration showed to have improved the winter discharge simulations as compared to the year-long or open water calibration approach. The winter system also performed quite well in the year-long simulations as well, which was contrary to the established belief that targeted calibration may worsen the year-long performance.

The fact that hydrological models cannot account for the complex river ice processes in the river could be a source of deviation between modeled and observed discharge. Events like rain on snow can also cause the models to generate streamflow, when in reality it has the potential to freeze in place instead of becoming runoff.

The three systems were deployed for forecasting purposes and their performance was evaluated against different performance metrices. The winter system yielded the lowest MCRPS score indicating low mean error in the probability sphere for forecasting, however, this system suffered from significant over-forecasting. The Year-long system performed adequately well for all metrices. The open water system displayed some tradeoffs in terms of MCRPS but in other metrices it also did well and performed closely with the Year-long model. However, the results show that the three systems had comparable performance, and neither outperformed the other significantly.

These preliminary findings are based on one winter forecasting season and do not provide sufficient insights into the performance of these systems. Moreover, the metrices used in this study are recognized to provide low details on forecast evaluation (Brown, Demargne et al. 2010) and other metrices such as rank histograms, reliability diagram, ROC etc. should be used to get more details on the performance of these systems. Future work will focus on further testing these systems through extended forecasting over multiple winter seasons and using more detailed forecast evaluation methods to harness the full potential of these systems.

5 Acknowledgments

This research project was funded by the Ministère de Sécurité publique (MSP) of the Quebec Government under the FLUTEIS project (project number CPS 18-19-26). We would like to thank Pascal Marceau, and Jean-Philippe Baril-Boyer the MSP for their continuous support and collaboration. The authors would like to acknowledge the support by Dominic Roussel from the Ministry of the Environment and the Fight Against Climate Change of Québec (MELCCFP) to provide relevant data as well as insights into MELCCFP operational system to carry out this study.

6 References

Beltaos, S. (2021). "Assessing the Frequency of Floods in Ice-Covered Rivers under a Changing Climate: Review of Methodology." Geosciences **11**(12): 514.

Bessar, M. A. (2021). Évaluation de l’incertitude liée à la modélisation hydraulique au sein d’un système de prévision d’ensemble des niveaux d’eau. PhD, Université Laval.

Boelee, L., D. M. Lumbroso, P. G. Samuels and H. L. Cloke (2019). "Estimation of uncertainty in flood forecasts—A comparison of methods." Journal of Flood Risk Management **12**(S1): e12516.

Boucher, M.-A., J. Quilty and J. Adamowski (2020). "Data Assimilation for Streamflow Forecasting Using Extreme Learning Machines and Multilayer Perceptrons." Water Resources Research **56**(6): e2019WR026226.

Brown, J. D., J. Demargne, D.-J. Seo and Y. Liu (2010). "The Ensemble Verification System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations." Environmental Modelling and Software **25**(7): 854-872.

Duan, Q., S. Sorooshian and V. Gupta (1992). "Effective and efficient global optimization for conceptual rainfall-runoff models." Water Resources Research **28**(4): 1015-1031.

Fortin, V., M. Abaza, F. Anctil and R. Turcotte (2014). "Why Should Ensemble Spread Match the RMSE of the Ensemble Mean?" Journal of Hydrometeorology **15**(4): 1708-1713.

Fulton, J. W., M. F. Henneberg, T. J. Mills, M. S. Kohn, B. Epstein, E. A. Hittle, W. C. Damschen, C. D. Laveau, J. M. Lambrecht and W. H. Farmer (2018). "Computing under-ice discharge: A proof-of-concept using hydroacoustics and the Probability Concept." Journal of Hydrology **562**: 733-748.

Gaborit, É., F. Anctil, V. Fortin and G. Pelletier (2013). "On the reliability of spatially disaggregated global ensemble rainfall forecasts." Hydrological Processes **27**(1): 45-56.

Hamilton, S., D. Hutchinson and R. Moore (2000). "Estimation of Winter Streamflow Using a Conceptual Hydrological Model: A Case Study, Wolf Creek, Yukon Territory." Journal of Cold Regions Engineering - J COLD REGIONS ENG **14**.

KlemeŠ, V. (1986). "Operational testing of hydrological simulation models." Hydrological Sciences Journal **31**(1): 13-24.

Kling, H., M. Fuchs and M. Paulin (2012). "Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios." Journal of Hydrology **424-425**: 264-277.

Kottek, M., J. Grieser, C. Beck, B. Rudolf and F. Rubel (2006). "World Map of the Köppen-Geiger Climate Classification Updated." Meteorologische Zeitschrift **15**: 259-263.

Levesque, É., F. Anctil, A. N. N. Van Griensven and N. Beauchamp (2008). "Evaluation of streamflow simulation by SWAT model for two small watersheds under snowmelt and rainfall." Hydrological Sciences Journal **53**(5): 961-976.

Liu, Y., A. H. Weerts, M. Clark, H. J. Hendricks Franssen, S. Kumar, H. Moradkhani, D. J. Seo, D. Schwanenberg, P. Smith, A. I. J. M. van Dijk, N. van Velzen, M. He, H. Lee, S. J. Noh, O. Rakovec and P. Restrepo (2012). "Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities." Hydrol. Earth Syst. Sci. **16**(10): 3863-3887.

Ministère de l'Environnement, d. l. L. c. l. C. C. (2023). "Normales climatiques du Québec 1981-2010." from https://www.environnement.gouv.qc.ca/climat/normales/sommaire.asp?cle=7027283.

Perrin, C. (2000). Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative.

Ricard, S., P. Lucas-Picher, A. Thiboult and F. Anctil (2022). Producing reliable hydrologic scenarios from raw climate model outputs without resorting to meteorological observations.

Seiller, G., F. Anctil and C. Perrin (2012). "Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions." Hydrol. Earth Syst. Sci. **16**(4): 1171-1189.

Singh, S. K. and A. Bárdossy (2012). "Calibration of hydrological models on hydrologically unusual events." Advances in Water Resources **38**: 81-91.

Teal, M. J., R. Ettema, J. F. Walker, P. Engel, Y. L. Lau, F. E. Hicks and P. M. Steffler (1996). "Estimation of Mean Flow Velocity in Ice-Covered Channels." Journal of hydraulic engineering. **122**(8): 474.

Thiboult, A. (2015). Toward coherent accounting of uncertainty in hydrometeorological modeling, Université Laval.

Thiboult, A. (2019). User Manual: HOOPLA version 1.0.2.

Thiboult, A. and F. Anctil (2015). "Assessment of a multimodel ensemble against an operational hydrological forecasting system." Canadian Water Resources Journal / Revue canadienne des ressources hydriques **40**(3): 272-284.

Thiboult, A. and F. Anctil (2015). "On the difficulty to optimally implement the Ensemble Kalman filter: An experiment based on many hydrological models and catchments." Journal of Hydrology **529**: 1147-1160.

Thiboult, A., F. Anctil and M. A. Boucher (2016). "Accounting for three sources of uncertainty in ensemble hydrological forecasting." Hydrology and Earth System Sciences **20**(5): 1809-1825.

Tolson, B. A. and C. A. Shoemaker (2007). "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration." Water Resources Research **43**(1).

Turcotte, R., A.-C. Favre, P. Lacombe, C. Poirier and J.-P. Villeneuve (2005). "Estimation des débits sous glace dans le sud du Québec : comparaison de modèles neuronal et déterministe." Canadian Journal of Civil Engineering **32**(6): 1039-1050.

Valdez, E. S., F. Anctil and M. H. Ramos (2022). "Choosing between post-processing precipitation forecasts or chaining several uncertainty quantification tools in hydrological forecasting systems." Hydrol. Earth Syst. Sci. **26**(1): 197-220.

Valéry, A., V. Andréassian and C. Perrin (2014). "'As simple as possible but not simpler': What is useful in a temperature-based snow-accounting routine? Part 2 – Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments." Journal of Hydrology **517**: 1176-1187.

Tables and Figures:

Table 1 Summary of KGEm Results for the three Systems.

Modelling System	Calibration (2008-2012)			Validation (2014-2018)					
				Year-long hydrograph			Winter specific hydrograph		
	Average	Min	Max	Average	Min	Max	Average	Min	Max
Winter system	0.92	0.88	0.93	0.92	0.85	0.94	0.88	0.78	0.92
Year-long system	0.92	0.91	0.94	0.94	0.91	0.96	0.85	0.74	0.92
Open water system	0.93	0.91	0.95	0.91	0.86	0.95	0.79	0.69	0.90

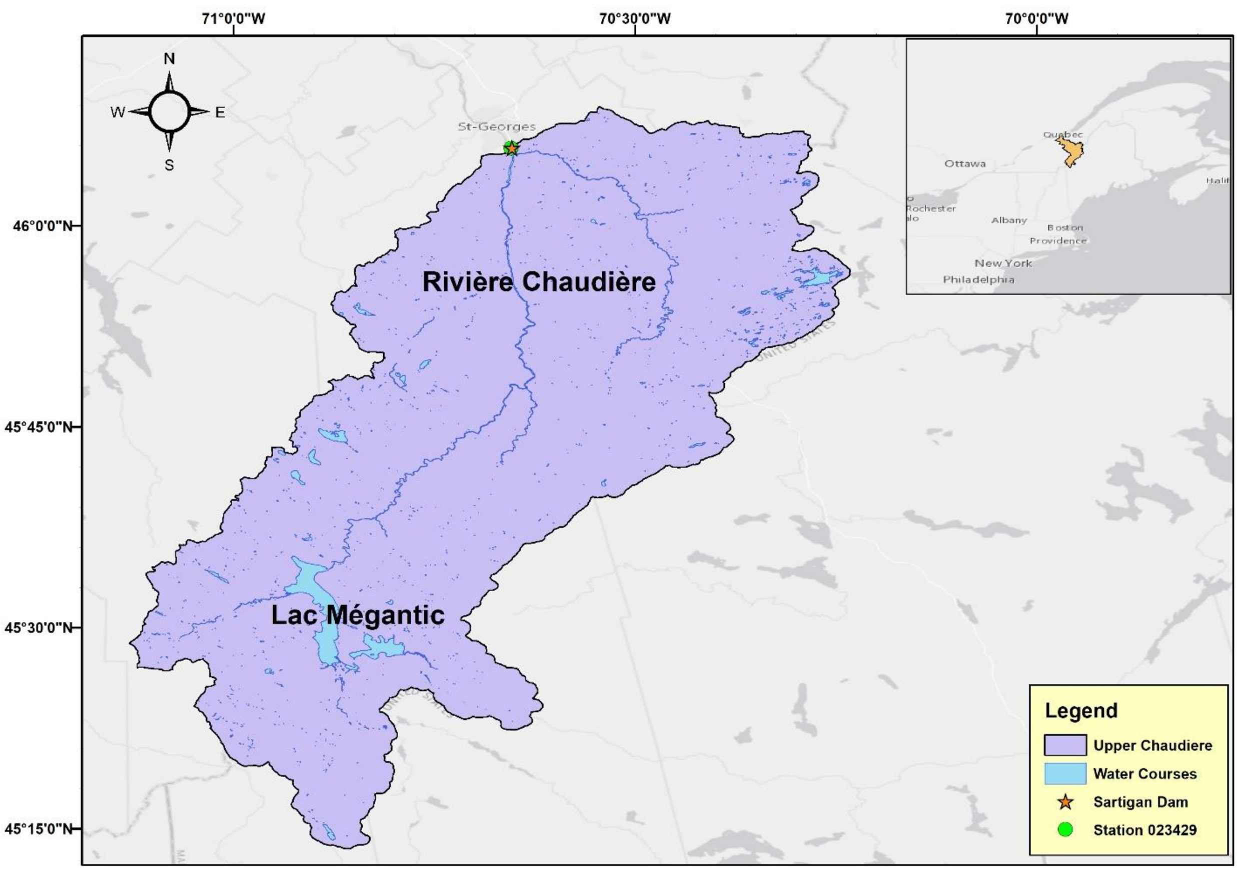

Figure 1 Location of the Upper Chaudière catchment used in the study.

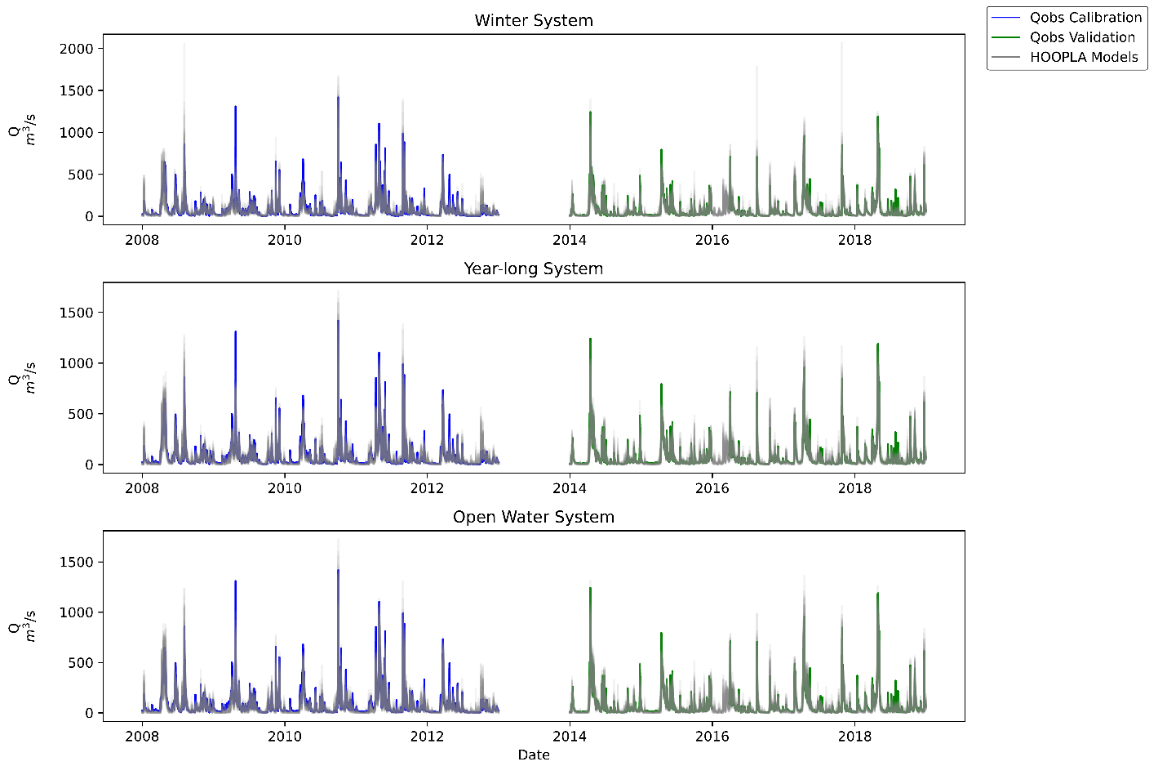

Figure 2 Calibration and Validation of the HOOPLA framework under the three calibration systems (the blue hydrograph represents the calibration period, the green hydrograph represents the validation period, grey hydrographs represent the HOOPLA framework).

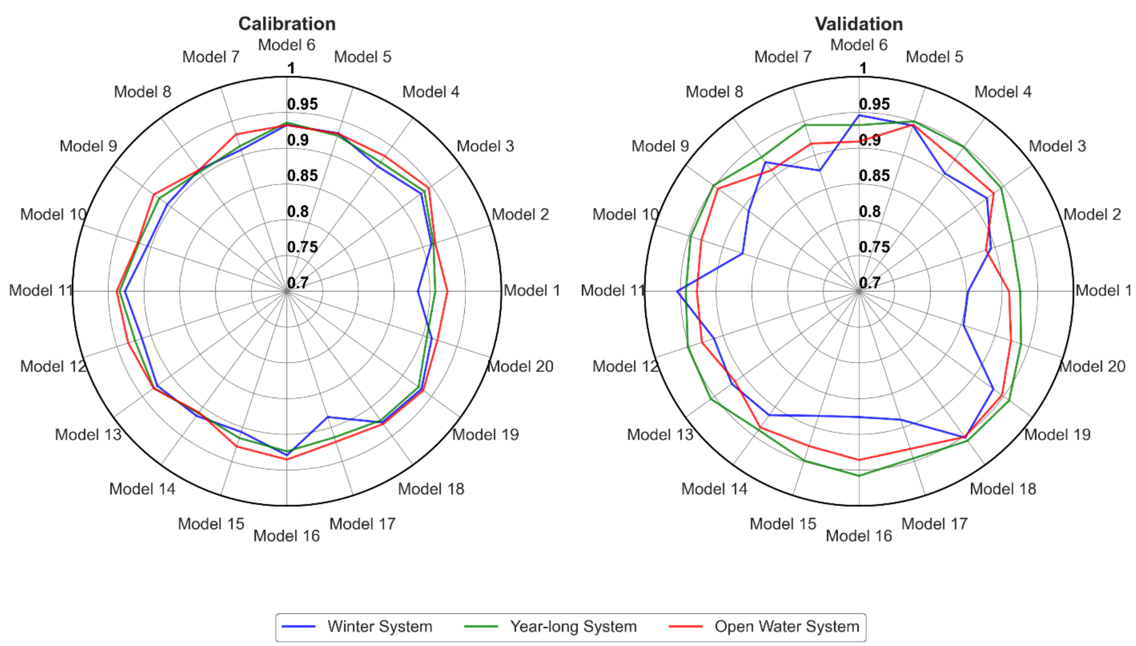

Figure 3 KGEm analysis for calibration and validation of Models in the HOOPLA framework.

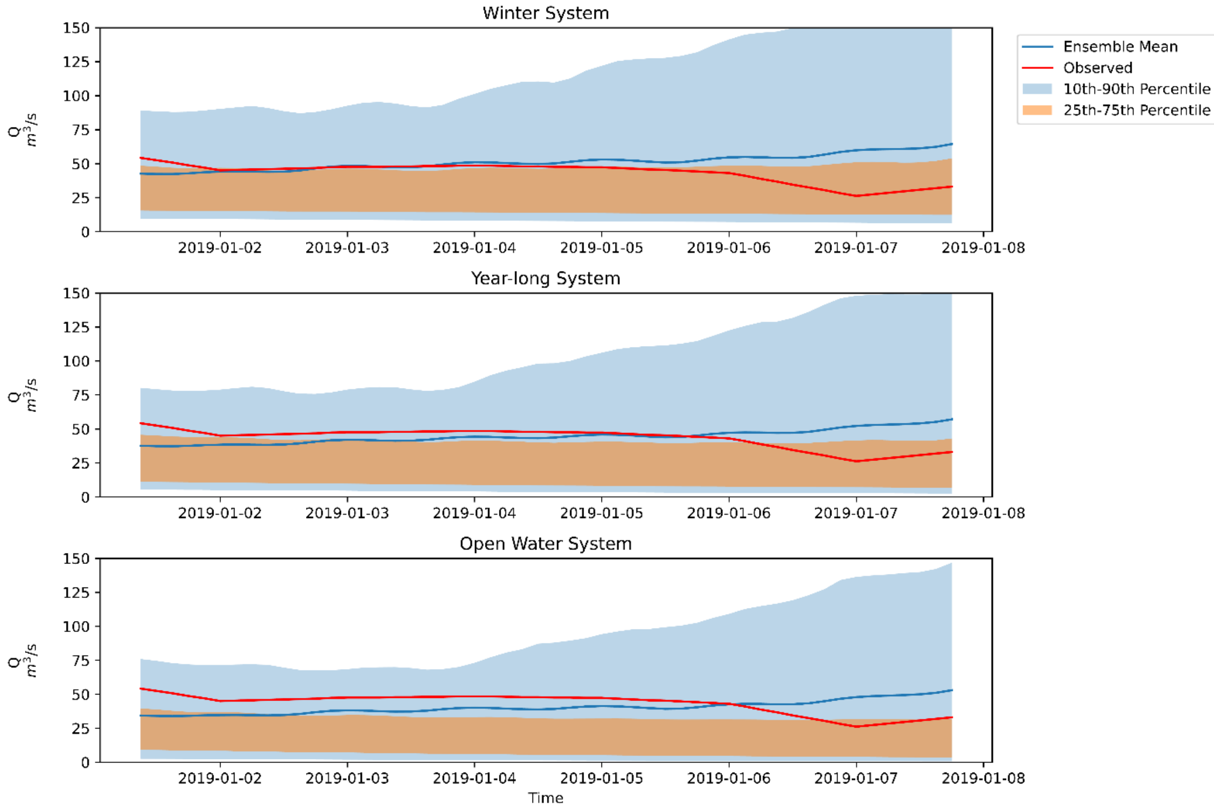

Figure 4 Single forecast produced by the three systems showing their performance and the forecast uncertainty.

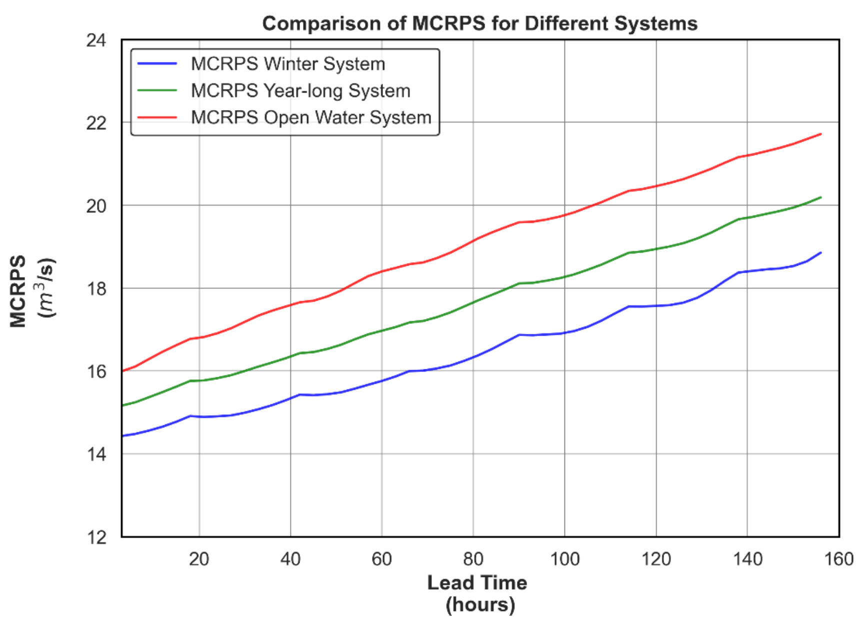

Figure 5 MCRPS comparison for the three systems.

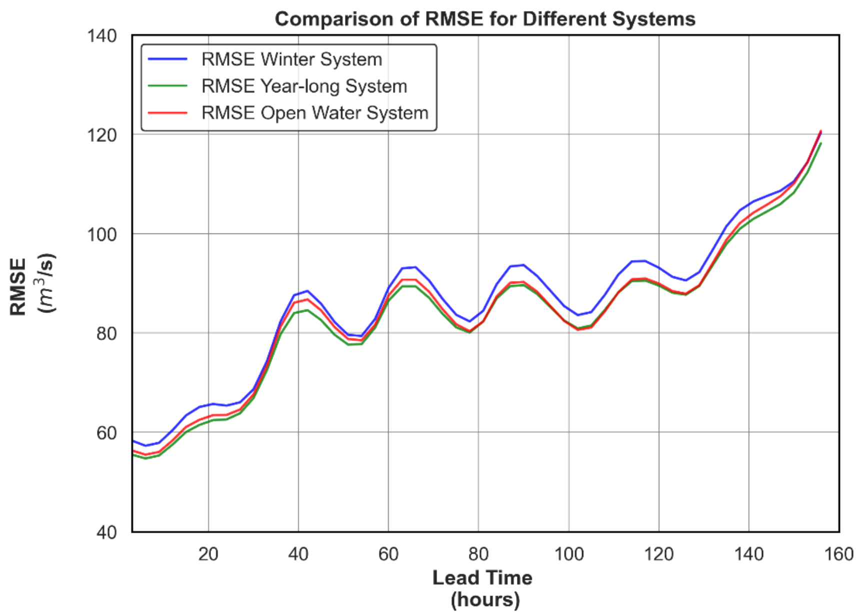

Figure 6 RMSE comparison for the three systems.

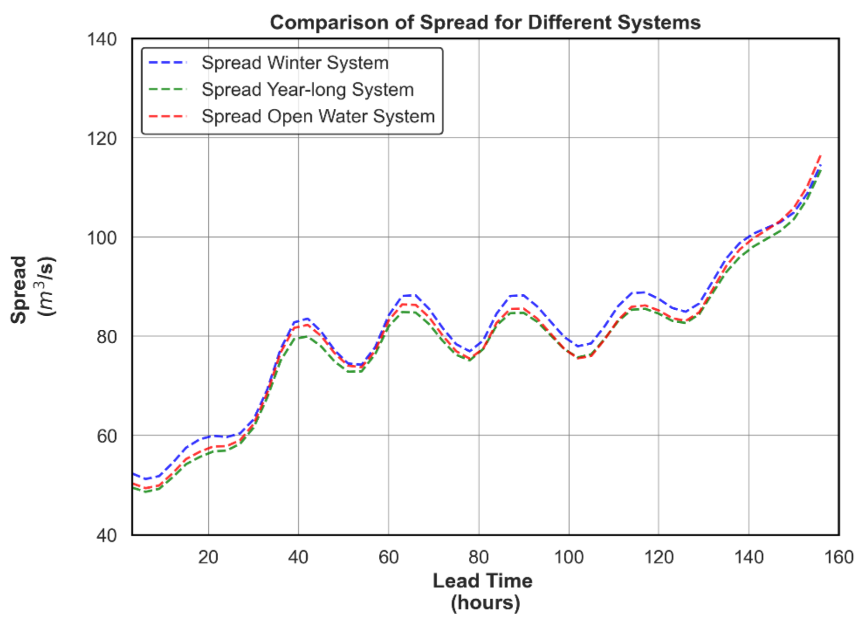

Figure 7 Comparison of ensemble spread for the different systems.

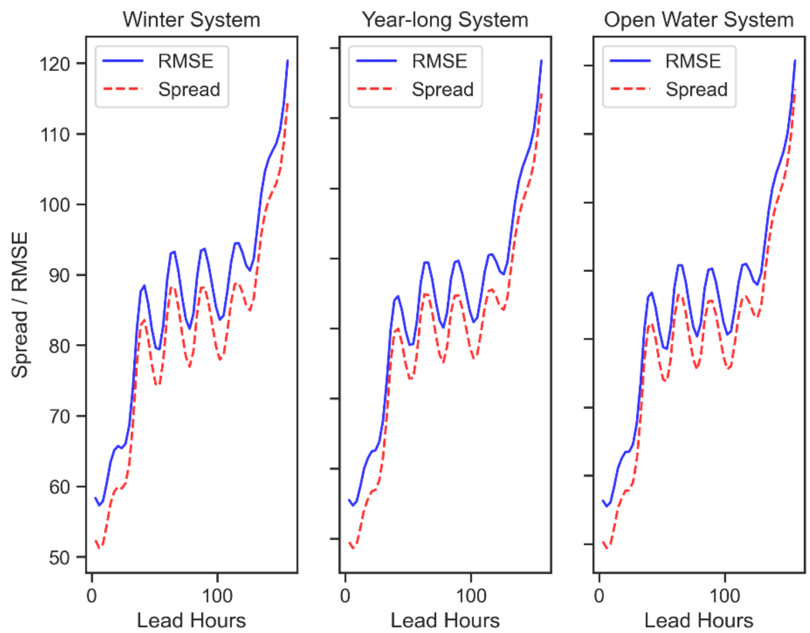

Figure 8 Reliability Evaluation through Spread-RMSE Plot.

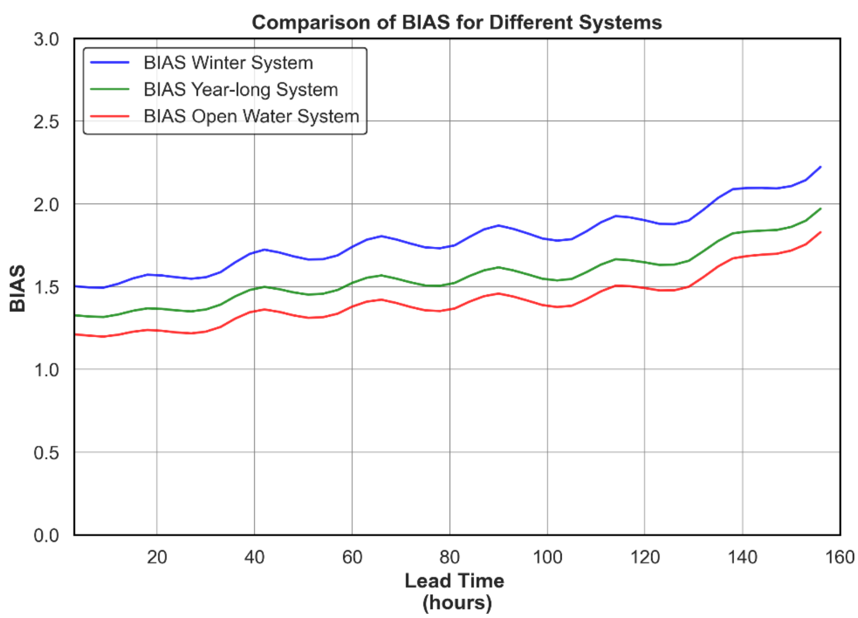

Figure 9 Comparison of the relative bias for the three systems.