Journal of Hydrology 642 (2024) 131861

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Application of weather post-processing methods for operational ensemble hydrological forecasting on multiple catchments in Canada

Freya Saima Aguilar Andrade [a,*], Richard Arsenault [a], Annie Poulin [a], Magali Troin [a,b], William Armstrong [a]

[a] Hydrology, Climate and Climate Change Laboratory, École de Technologie Supérieure, 1100 Notre-Dame W., Montréal (Quebec), Canada
[b] HydroClimat TVT, Maison du Numérique et de l'Innovation, Place Georges Pompidou, 83 000 Toulon, France

ARTICLE INFO

Keywords:
Hydrological forecasting
Weather post-processing
Quantile mapping
Ensemble forecasting
Forecast verification

ABSTRACT

Hydrological forecasts contain biases that need to be addressed for their effective use in operational decision-making in water resources management. Performing post-processing allows reducing the overall systematic bias while improving the distribution and accuracy of hydrological forecasts. In this study, a Quantile Mapping (QM) post-processing method was applied on weather forecasts following three temporal configurations (monthly, seasonal, and annual) of the quantile mapping scheme. The evaluation encompasses 20 catchments in southern Canada, employing a leave-one-out approach with the QM method on ECMWF ensemble weather forecasts spanning 2015–2020 inclusively. These processed forecasts are subsequently utilized as forcings for eight hydrological models, generating ensemble streamflow forecasts over a 6-year period with a lead time of 10 days and a sub-daily timestep of 6 h. The performance of the QM method is mainly assessed using the Continuous Ranked Probability Score (CRPS) metric, in complement with a forecast reliability score (ABDU) and a forecast sharpness metric (NMIQR). Significant improvements are discerned in precipitation forecasts upon the application of QM. Notably, these improvements are translated into enhanced hydrological forecasts for over half of the catchments studied (55 %). Surprisingly, no discernible differences in performance are observed among the three QM configurations in most catchments. Interestingly, there are watersheds where the implementation of QM exhibit either poorer or no change in performance and sharpness compared to raw forecasts.

1. Introduction

Hydrological forecasts play a crucial role in the operational water resources sector. For hydropower and water resource management, skillful and reliable streamflow forecasts constitute a critical input for the operation and management of infrastructures (Pagano et al., 2014). The hydrological cycle of many regions worldwide, including Canada, is dominated by snow accumulation and ablation processes. Most floods in these regions result from a combination of factors, including saturated or frozen soils, heavy rainfall, and rapid melting of snow within a 24-hour period, facilitated by high winds and strong thermal and moisture advection (Graybeal and Leathers, 2006; Pomeroy et al., 2016). Also flooding events often result from intense sub-daily rainfall (Dale, 2021), with convective rainfall becoming more frequent and intense due to climate change (Berg and Haerter, 2013). In a study by Guerreiro et al. (2018) it was found that extreme sub-daily precipitation exhibited a stronger response to climate warming compared to extreme daily precipitation. Therefore, it is essential to conduct sub-daily hydrological forecasts as they play a crucial role in operational forecasting, particularly in predicting floods.

Ensemble forecasting systems based on numerical weather prediction (NWP)-forced hydrological models are one suitable approach to simulate and predict river flow (Troin et al., 2021). NWP models use estimates and assumptions to predict future weather. However, they have some limitations related their structure and parametrization. Such limitations in NWP models can result in less accurate weather forecasts (Rayner et al., 2005; Wu et al., 2011; Robertson et al., 2013; Tao et al., 2014) which can propagate errors into the forecasting system and affect the accuracy of streamflow forecasts (Leutbecher & Palmer, 2008). Uncertainty in ensemble streamflow forecasting systems can be reduced by post-processing. Statistical ensemble post-processing methods are used to compensate for errors in model structure, and to correct systematic biases that may be present in the raw ensemble forecasts (Madadgar et al., 2014; Hopson et al., 2018). Many studies have shown

* Corresponding author.
E-mail address: freya-saima.aguilar-andrade.1@ens.etsmtl.ca (F.S. Aguilar Andrade).

https://doi.org/10.1016/j.jhydrol.2024.131861
Received 8 September 2023; Received in revised form 4 July 2024; Accepted 11 August 2024
Available online 18 August 2024

that statistical post-processing is able to improve weather forecast accuracy. Weather post-processing methods are applied to adjust the mean and variance of the forecasts to better match the distribution of the observed data (Verkade et al., 2013). As the hydrological response is sensitive to climate variability, an accurate post-processing is needed to represent and preserve the space–time covariance of the weather patterns over a given catchment (Hopson et al., 2018).

Quantile mapping (QM) is the most common approach for post-processing weather forecasts (e.g., Fang et al., 2015; Ghimire et al., 2019; Niranjan et al., 2022; Verkade et al., 2013). It involves adjusting the raw weather forecast to match the distribution of the observations. QM improves the accuracy of the forecast by making it more consistent to historical observations. It is among the more flexible bias correction methods that attempt to adjust the variance of the model distribution to better match the observed variance (Maraun, 2016). An example is done by Baker et al. (2019), where QM successfully removes the systematic bias in the CFSv2 reforecasts. In another study, Themeßl et al. (2011) compared various downscaling and error correction methods and showed that QM method performs best for daily precipitation. Wilcke et al. (2013) evaluated the QM post-processing technique for bias correction in four weather variables: temperature, precipitation, relative humidity, and wind speed. They showed that annual and monthly biases are reduced to close to zero by QM for all variables, with the daily precipitation as the variable that was strongly improved. Precipitation is largely considered as the input data with the largest source of uncertainty (Biemans, et al., 2009; McMillan et al., 2011; McMillan et al., 2012).

Errors in initial model conditions at the time of forecast can lead to biases between the observed and simulated streamflow, rendering forecasts unrealistic for the first lead time. These errors can be reduced by data assimilation methods which aim at updating the model state variables or parameters through assimilation of observations. Data assimilation is used in hydrological modelling and forecasting with a variety of methods and assimilated data types, such as streamflow, snow water equivalent, and soil moisture (Liu et al., 2012; Zhang, 2015). Bourgin et al. (2014) investigated the role of data assimilation and post-processing of streamflow to enhance the skill of a hydrological ensemble forecasting system. They found that data assimilation significantly improved the quality of the ensemble mean, while post-processing markedly enhanced the reliability of streamflow forecasts. Using both methods together led to more reliable and sharp streamflow forecasts, although the impact on forecast reliability is stronger from post-processing. However, the study was limited by the use of a single data assimilation technique, one post-processing method, and one lumped conceptual model to generate streamflow forecasts. The authors recommend comparing different post-processing methods and alternative hydrological models across multiple catchments to reach more general conclusions. Following this recommendation, our intention is to utilize an ensemble of hydrological models and investigate the uncertainty arising specifically from forecasted precipitation.

The objective of this study is to improve the accuracy and reliability of short-term (hourly and sub-daily) streamflow forecasts from small to medium-sized catchments in Canada. For this, the Quantile Mapping post-processing method is applied to the European Centre for Medium-Range Weather Forecasts (ECMWF). We assess this technique in using multi-year hindcasts and eight different hydrological models with data assimilation in a large sample of catchments. Through this study, attempts are made to understand the limitations of post-processed precipitation based on catchment attributes and hydrological model structure. The next section describes the study area and methods implemented to investigate this issue. Section 3 presents the obtained results, which are discussed in Section 4. Concluding remarks are provided in Section 5.

2. Materials and methods

2.1. Study area

Twenty catchments were selected with different surface areas from 30 km^2 to up to 35 000 km^2 (Table 1). This selection was based on the availability of observed streamflow data over at least a period of 10 to 15 years at a 6-hourly time-step to meet the objective of this study. More catchments were initially considered in this study, however, since streamflow data at the sub daily scale is not readily available in Canada and must be requested manually (except in the province of Quebec), the datasets that were made available were limited in spatial distribution. Among other things, most hydrometric gauges with data at the hourly or sub-daily scales were regulated rivers, which made modelling and forecasting problematic. The final selected catchments are mainly located in eastern Canada (Fig. 1). A comprehensive description of the catchments' characteristics can be found in Table 1.

2.2. Hydrometeorological observation and weather forecast data

The meteorological input data necessary for the eight hydrological models include the minimum and maximum air temperatures, as well as the total precipitation, captured at 6-hour intervals. The historical data come from the ERA5 reanalysis database (Hersbach et al., 2020). ERA5 is the fifth generation of global atmospheric reanalysis published by the European Centre for Medium-Range Weather Forecasts (ECMWF) that employs the assimilation of four-dimensional variable data (4DVar) from satellite and in-situ observations (Copernicus Products, 2018). The ERA5 dataset offers a spatial resolution of 0.25 degrees (~31 km at the study site) covering the entire globe, providing hourly data for a comprehensive range of variables encompassing oceanic, terrestrial, and atmospheric domains. It has been shown to be a reliable proxy for observed weather data, especially in regions with low weather station density such as in Canada (Tarek et al., 2020). This reanalysis data covered the 1981–2022 period, inclusively.

In this study, the decision to utilize ERA5 reanalysis data instead of observations for precipitation forecasts stems from several considerations. Firstly, the inherent limitations of using observations, including missing data at sub-daily time scales and spatial–temporal heterogeneity, pose significant challenges for post-processing. Moreover, many catchments lack sufficient high-quality observational records, further complicating the analysis. By leveraging ERA5 data, these issues are circumvented, allowing for a more focused approach to post-processing. Furthermore, although ERA5 reanalysis may contain biases inherent to its assimilation and modelling processes, using it as the reference for comparing precipitation forecasts minimizes the impact of these biases on post-processing.

Ensemble weather forecasting data was provided by the ECMWF's operational Integrated Forecasting System (IFS) which provided minimum and maximum air temperature as well as precipitation forecasts on a 6-hourly timestep over the entire domain. The ensemble forecasts are composed of 50 members at a lead-time of 10 days with a horizontal resolution of around 18 km. These forecasts were available for 2016–2020, inclusively. It is important to note that also ERA5 reanalysis data is derived directly from the ECMWF integrated forecasting system cycle 41r2, which has been operational since 2016. This integration ensures that the biases between forecasts and reference observations are not affected by the forecasting system initialization.

A minimum of a 10 to 15-year period of streamflow observations is required to perform a reliable model calibration (Yapo et al., 1996). The period of 10 to 15 years needs to include the forecasting 2015–2020 period to train the QM post-processing methods and to test it on an independent period. The observed daily hydrometric data was obtained from the Water Survey of Canada website and covers the period 1997–2022. To align with the project requirements of a 6-hour time-step, a request was made to Environment and Climate Change Canada

Table 1
Main characteristics of the 20 catchments used in this study. The ordering of the catchments follows their location in the provinces of Canada from west to east.

	ID	Station Name/ Catchment	Province*	Drainage area (km^2)	Lat.	Lon.	Average total annual prec. (mm)	Average annual runoff (mm)
1	07JD002	Wabasca River At Highway No. 88	AB	35,800	57.87	−115.39	512	65
2	02GA010	Nith River Near Canning	ON	1030	43.19	−80.46	1003	388
3	40,406	Petite Nation	QC	1331	46.13	−75.13	1074	577
4	40,624	Du Lièvre	QC	4560	47.25	−74.86	1064	673
5	40,204	Rouge	QC	5479	46.40	−74.67	1121	719
6	40,110	Du Nord	QC	1163	46.01	−74.22	1165	796
7	52,219	L'assomption	QC	1286	46.35	−73.79	1058	712
8	24,014	Bécancour	QC	2163	46.19	−71.56	1254	801
9	23,401	Beaurivage	QC	708	46.47	−71.29	1239	721
10	23,402	Chaudière	QC	5820	45.98	−70.74	1177	735
11	23,303	Etchemin	QC	1152	46.53	−70.68	1275	862
12	23,106	Du Sud	QC	821	46.74	−70.51	1235	896
13	23,422	Famine	QC	696	46.26	−70.45	1147	798
14	01EF001	Lahave River At West Northfield	NS	1250	44.45	−64.59	1236	972
15	01CA003	Carruthers Brook Near St. Anthony	PE	46.8	46.74	−64.19	1161	754
16	01EO001	St. Marys River At Stillwater	NS	1350	45.17	−61.98	1262	1104
17	03OE011	Pinus River	NL	780	53.15	−61.56	1000	771
18	02XA003	Little Mecatina River Above Lac Fourmont	NL	4540	52.23	−61.32	1040	725
19	03QC002	Alexis River Near Port Hope Simpson	NL	2310	52.65	−56.87	1055	819
20	02YA002	Bartletts River Near St. Anthony	NL	33.6	51.45	−55.64	1170	712

*Provinces internationally approved alpha code: AB (Alberta), ON (Ontario), QC (Quebec), NS (Nova Scotia), PE (Prince Edward Island) and NL (Newfoundland and Labrador). All hydrometric stations can be found in the site of Environment and Climate Change Canada and the Direction de l'Expertise Hydrique of the province of Quebec.

to acquire the data in their native 5-minute time-steps, after which the data were processed at the 6-hour intervals.

While acknowledging potential biases between precipitation and streamflow datasets, our study adopts a comprehensive approach to address these challenges. This includes detailed hydrological model calibration and streamflow post-processing, which will be elaborated on in the subsequent sections.

2.3. Hydrological models

This study was conducted with eight lumped hydrological models, CEQUEAU (Girard et al., 1972), a sub-daily version of GR4J (we coined GR5dt) (Perrin et al., 2003; Mathevet, 2005), HBV (Bergström & Forsman, 1973), HYMOD (Wagener et al., 2001), IHACRES (Jakeman et al., 1990), MOHYSE (Fortin & Turcotte, 2007), SIMHYD (Chiew et al., 2002), and TOPMODEL (Beven et al., 1984), which are briefly described below and summarized in Table 2. All of these hydrological models are widely accessible; however, certain modifications have been implemented in their configurations to enable their utilization within a consistent lumped conceptual framework. All hydrological models are applied in the exact same framework: 6-hour time-step computations with the same input data: precipitation and mean temperature. Furthermore, all models share the same snow accumulation and melt model and the same potential evapotranspiration (PET) model. In this case, the snow module used is the CemaNeige module (Valéry, 2010) and the PET formulation is that of Oudin et al. (2005a, 2005b). CemaNeige is a degree-day snow module which simulates the snowpack dynamics by using two parameters: the snow cover condition coefficient and the amount of snow melting related to the air temperature (mm/^{o}C). CemaNeige was adapted to Canada by simplifying it to a single altitude band and modified to run at the sub-daily time step. Oudin's PET formulation is expressed by equation (1):

$$PET = 1000\frac{\mathrm{R}_e}{\lambda\rho}\frac{\mathrm{T}_m + 5}{100} \tag{1}$$

In this equation, $\mathrm{T_m}$ is the mean daily air temperature (°C); λ is the slope of the vapor pressure curve (kPa/°C), calculated using $\mathrm{T_m}$; ρ is the density of water (kg/L) and $\mathrm{R_e}$ is the extraterrestrial solar radiation in MJ/($\mathrm{m}^2\cdot$ day), which is determined based on the date and latitude. The PET model is executed at the sub-daily time step. This finer temporal resolution is employed to capture specific hydrological processes more accurately, notably snowmelt and evapotranspiration, both of which are significantly influenced by the diurnal cycle. An overview of each model is given below, and a summary is presented in Table 2.

CEQUEAU is a semi-distributed hydrological model that divides the catchment into predetermined grid cells of equal area. It consists of two components: a rainfall-runoff conceptual module that simulates the hydrological states of the catchment based on precipitation, air temperature, and physiographic inputs, and a transfer function that routes water downstream to simulate discharge. The model incorporates various reservoirs representing snowpack, evapotranspiration, unsaturated and saturated zone water, and storage in lakes and marshes. The model enables the vertical distribution of water and updates the state of these reservoirs, allowing for the simulation of the hydrological processes in the catchment. Some changes have been made in the CEQUEAU structure in order to use it in a homogeneous lumped conceptual framework.

GR5dt is a rainfall-streamflow model adapted for multiple time steps (hourly to daily time steps) (Ficchi, 2017; Ficchi et al., 2019). The model is divided into two stores: a production store and a routing store. GR5dt is based on the combination of the lumped GR4J and GR4H models (Génie Rural à 4 paramètres Journalier; Perrin et al. 2003, Mathevet 2005) with five free parameters to calibrate: GR5dt uses the four GR4J parameters and the GR4H parameter that describes the time base in hours at the unit hydrograph.

HBV (Hydrologiska Byrans Vattenavdelning) is a hydrological model that simulates streamflow using rainfall, temperature, and estimated evaporation as input data. It includes modules for soil water, evaporation, and groundwater, which is described by three linear reservoirs. Channel routing is done using a triangular weighting function.

HYMOD (HYdrological MODel; Wagener et al., 2001) is a rainfall excess model based on a nonlinear water storage capacity distribution function. The water stemming from rainfall and snowmelt is first infiltrated to the unsaturated zone, where evaporation is determined based on soil moisture. Streamflow is generated based on the spatial distribution of the catchment's maximum storage capacities. The streamflow is divided into "quick" (surface) and "slow" (baseflow) flows, which eventually determine the final streamflow.

IHACRES (Identification of unit Hydrographs And Component flows

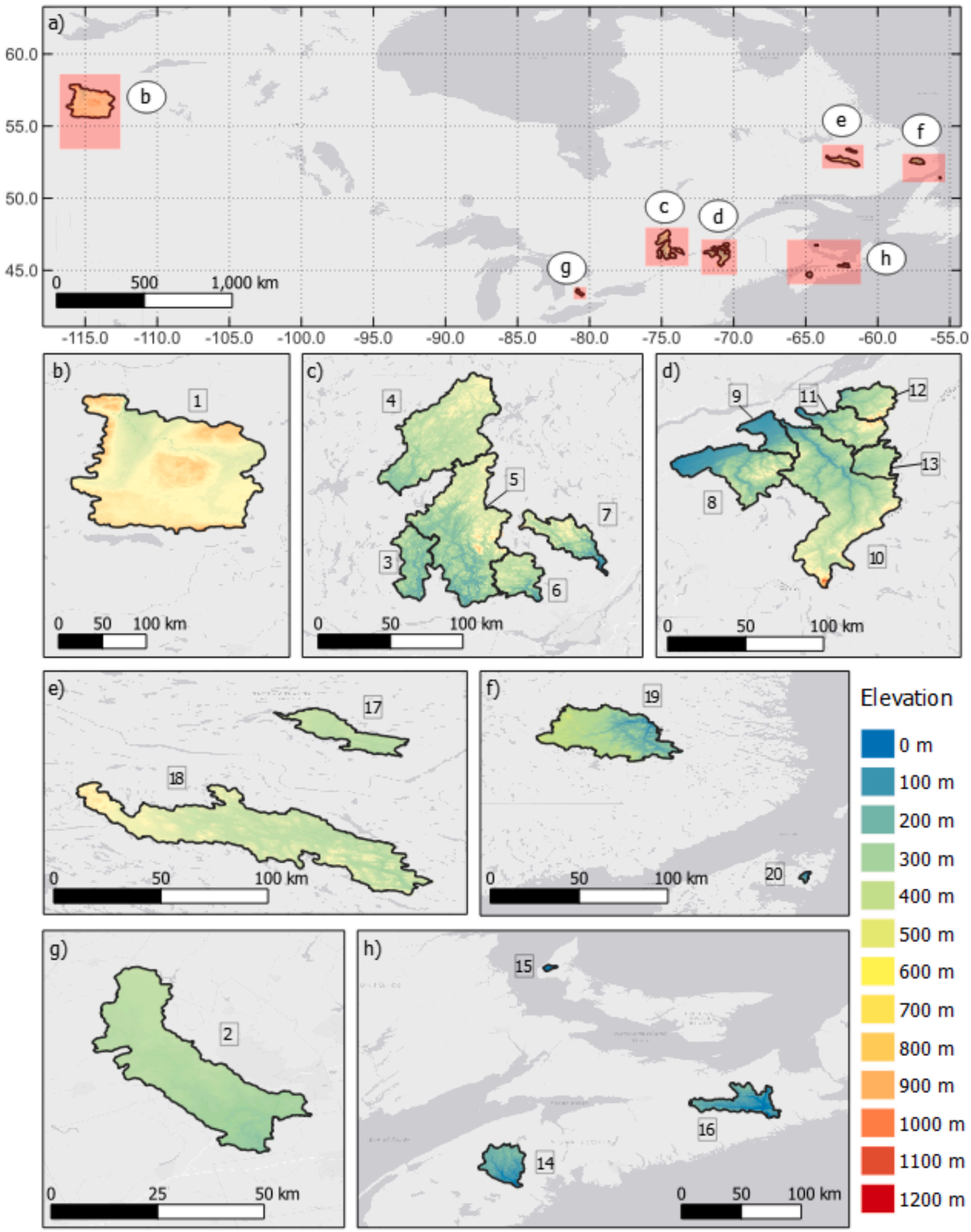

Fig. 1. Location of the 20 catchments used in this study, including their elevation profiles. Numbers alongside each catchment represent their number in Table 1 for reference.

from Rainfall, Evapotranspiration and Streamflow) is a conceptual rainfall-streamflow model. The model characterizes catchment-scale hydrological behaviour with six parameters. The rainfall-streamflow processes are represented by two modules: a non-linear loss module and a linear routing module.

MOHYSE (Modèle Hydrologique Simplifié à l'Extrême) is a lumped and conceptual model that operates at the daily and sub-daily time step (Fortin and Turcotte, 2007). MOHYSE comprises two compartments: a

Table 2
Main characteristics of the eight lumped hydrological models used in this study.

Name	Number of calibrated parameters	Number of storages	Derived from
CEQUEAU	9	2	Girard et al. (1972)
GR5dt	5	2	Perrin et al. (2003) & Mathevet (2005)
HBV	9	3	Bergström and Forsman (1973)
HYMOD	6	5	Wagener et al. (2001)
IHACRES	7	3	Jakeman et al. (1990)
MOHYSE	10	2	Fortin and Turcotte (2007)
SIMHYD	8	3	Chiew et al. (2002)
TOPMODEL	7	3	Beven et al. (1984)

vadose zone production store and an aquifer routing store and encompasses a set of 10 calibration parameters. The input climate variables are precipitation and mean temperature. In MOHYSE, precipitation and melt water are added before computing the evaporation, infiltration and routing towards the stream A unit hydrograph is employed to aggregate and rout the distinct water fluxes generated through surface, vadose zone, and aquifer interactions to the outlet.

SIMHYD is a conceptual rainfall-streamflow model that estimates daily and sub-daily streamflow from precipitation and areal potential evapotranspiration data. The model accounts for several key processes including infiltration, evapotranspiration, soil moisture, and groundwater for generating streamflow (Chiew et al., 2002). It uses a one-layer evaporation model and considers both infiltration-excess and saturation-excess streamflows by using an interception store and a soil moisture store.

TOPMODEL (TOPographic MODEL) is a semi-distributed conceptual model (Beven et al., 1984). Total streamflow is calculated as the sum of two components: saturation excess overland flow from variable contributing areas, and subsurface flow from the saturated zone of the soil. The model is based on three assumptions: (1) a sequence of steady state representations can approximate the behavior of the saturated zone, (2) the slope of the ground surface is a reasonable estimate of the hydraulic gradient, and (3) the transmissivity of the downslope area can be expressed as an exponential function of the storage deficit or water table depth. These assumptions result in a simple relationship between the catchment storage (or storage deficit) and the local water table level (or drainage-caused storage deficit).

The hydrological models were chosen because of their successful implementation and efficient performance in Canadian catchments (Dion et al., 2021).

2.4. Model calibration

The eight hydrological models were calibrated with the Covariance Matrix Adaptation Evolution Strategy (CMAES) optimization algorithm with 5,000 evaluations per model (Hansen and Ostermeier, 1997; 2001). CMAES is an effective algorithm for optimizing hydrological model parameters (Arsenault et al., 2014). The calibration of the hydrological models was performed on the 1997–2022 period, corresponding to the intersection of available meteorological and hydrometric data. Model calibration was performed using the using the Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970) as the objective function. The NSE ranges betwen $-\infty$ and 1 with NSE=1 being the optimal value. It is expressed as follows:

$$NSE = 1 - \left[\frac{\sum_{t=1}^{n}\left(Q_t^{obs} - Q_t^{sim}\right)^2}{\sum_{t=1}^{n}\left(Q_t^{obs} - Q^{mean}\right)^2}\right] \tag{2}$$

where Q_t^{obs} and Q_t^{sim} are the observed and simulated streamflow at the 6-hour time-step, respectively; Q^{mean} is the mean of observed streamflows; and n is the number of observed or simulated streamflow values. The evaluation criteria are interpreted according to Moriasi et al. (2015) for a daily, monthly or annual hydrological analysis (flow), i.e. 0.50 < NSE<=0.70 is satisfactory, 0.70 < NSE<=0.80 is considered good and anything above is considered very good. Values below 0.50 are considered unsatisfactory.

No validation was performed, and all available data was used for calibration following best practices to maximize parameter information content as recommended by Arsenault et al. (2018), Shen et al. (2022), and Mai (2023). Instead, performance in forecasting is used to assess the calibrated model performance.

2.5. Data assimilation of initial hydrological model states

Data assimilation was performed with the Ensemble Kalman Filter (EnKF) which uses perturbed meteorological forcings to generate a set of probable states that are updated based on the model similarity to observed flows for nonlinear filtering issues (Evensen, 2003). The EnKF is extensively used in hydrological sciences and forecasting applications (Piazzi et al., 2021; Bergeron et al., 2021). EnKF generates an ensemble of probable initial states, allowing to quantify their uncertainty. The adjustment in initial model states is performed so that the distribution of state values matches the actual observations. Forecasts are then performed from each initial state, including the initial condition uncertainty in the forecasting process, which is another key advantage of EnKF. In this study, three hydrologic states were modified (i.e., water content in groundwater storage, vadose zone storage and snowpack water equivalent) to give maximum flexibility to the models. Furthermore, instead of using the full ensemble of probable initial states as generated by the EnKF, the mean ensemble is considered as the best estimator of the actual initial state. This was done to avoid combining the 50 weather forecasts to the 25 members of initial states and thus generating ensemble of 1250 forecast members, as was done in Dion et al. (2021). More details on EnKF can be found in Evensen (2003).

The use of both data assimilation and post-processing of precipitation in hydrological forecasting is highly recommended since data assimilation has a strong impact on forecast accuracy while post-processing affects forecast reliability (Bourgin et al., 2014; Dion et al., 2021).

2.6. Post-processing

In this study, weather forecasts, and specifically precipitation forecasts, were post-processed by using the QM method (Eq. (3). QM employs a quantile-based transformation of distributions: a quantile of the present day simulated distribution is replaced by the same quantile of the present-day observed distribution (Maraun, 2016).

$$x_{i,corr}^{f} = qD_y^p\left(pD_x^p\left(x_{i,raw}^{f}\right)\right) \tag{3}$$

Where the given time series is denoted as x_i, and future simulations and derived measures are indicated with a superscript *f*. The quantile for a probability α of a distribution D is represented as $qD(\alpha)$ and is defined as the value which is exceeded with a probability 1 − α when sampling from the distribution. The probabilities corresponding to a given quantile $qD(\alpha)$ (i.e. the cumulative distribution function CDF) are written as $pD(q) = \alpha$.

For instance, considering a watershed with 1817 days of calibration data (5 years) and 365 days of validation data, each day includes a 10-day precipitation forecast with a 6-hour time step, comprising 50 ensemble members. Consequently, each day's data can be represented as a [40x50] matrix. For each ensemble of members at each time-step for the entire calibration period, (which is a [1817x50] matrix reshaped to a vector of length 90850), 100 quantiles are calculated, for both observed and forecasted precipitation. This results in two [100x1] matrices: one for observed quantiles (qD_{obs}) and one for simulated quantiles (qD_{cal}), for each of the 40 lead-times. The correction factor ($quantile_{corr}$) for each quantile is then determined using the following equation:

$$quantile_{corr} = \frac{qD_{obs}}{qD_{cal}}$$

This process is repeated for all 40 time-steps, resulting in a [100x40] matrix of correction factors. Thus, there are 4000 correction factors. Given the probabilistic nature of precipitation forecasts—comprising 50 members—the correction factors are systematically applied across the ensemble. The member-quantile assignation is performed as: (i) if the member's precipitation is greater than the last quantile, it is assigned to the 100th quantile, (ii) if the member's precipitation matches an existing quantile, it is assigned accordingly and (iii) if the member's precipitation does not exactly match a quantile, it is assigned to the next higher quantile. This collective correction mechanism preserves the inherent variability within the forecast ensemble, contributing to a more realistic representation of precipitation scenarios.

The quantile mapping correction is performed by multiplying the member's raw precipitation value by the corresponding correction factor:

$$x_{valcorr} = x_{val} \times quantile_{corr}$$

where $x_{valcorr}$ is the corrected forecast member value, x_{val} is the raw forecast member in the validation period, and $quantile_{corr}$ is the quantile function probability of the observed distribution and simulated calibration period.

The QM methodology employed for the postprocessing of precipitation forecasts, involves a quantile-wise correction factor to the forecasted precipitation distribution, aligning it with the actual observed distribution. The correction factor is calculated independently for each quantile forecast, thereby ensuring an accurate adjustment of the entire forecast distribution.

Several studies applying quantile-based post-processing methods (including QM) have observed a general improvement in precipitation forecasts, particularly in correcting extreme rainfall events (Shastri et al., 2017; Tani and Gobiet, 2019; Niranjan Kumar et al., 2022). Although there is a wide variety of methods that can be used to correct bias in meteorological and hydrological forecasts (Li et al., 2017; Troin et al., 2021), in this study, we decided to use the Quantile Mapping (QM) correction method. This decision was based on its frequent mention in the scientific community as the most popular technique for precipitation forecast bias correction (e.g., Katiraie-Boroujerdy et al., 2020; Li et al., 2023; Maurer and Pierce, 2014; Miao et al., 2016; Trinh-Tuan et al., 2019; etc.).

To extend the evaluation period beyond the available 2015–2020 precipitation forecast data, a leave-one-out approach is adopted. This technique involves resampling one year for evaluation purposes and five years for training, thus creating an augmented dataset that spans a more comprehensive six-year timeframe. This prolonged evaluation period enhances the robustness of the methodology, allowing for a nuanced examination of its efficacy over an extended temporal domain.

The QM method was used in three different temporal configurations which are briefly described below:

1. Monthly QM: this configuration involves applying the QM method to precipitation forecasts pooled by month. The monthly distribution of the raw precipitation forecasts for each lead time is adjusted to match the monthly distribution of the observed precipitation.
2. Seasonal QM: this configuration involves applying the QM method to precipitation forecasts pooled by season. The seasonal distribution of raw precipitation forecasts for each lead time is adjusted to match the seasonal distribution of the observed precipitation (winter: DJF, spring: MAM, summer: JJA, and fall: SON).
3. Annual QM: this configuration involves applying the QM method to all precipitation forecasts at once. The annual distribution of the raw precipitation forecasts for each lead time is adjusted to match the annual distribution of the observed precipitation.

By applying the QM for different temporal configurations (i.e., monthly, seasonal, and annual) it is possible to evaluate the performance of the post-processing method at different temporal scales and to determine which configuration is the most appropriate for hydrological forecasting. Indeed, pooling forecasted and observed precipitation data on short periods (e.g., by months or seasons) means that their distributions will be more consistent with the underlying hydrological processes. However, pooling data in more (but smaller) pools also has the side effect of reducing the number of training points for the QM application, impacting thus the quality of the distributions. Therefore, by using multiple pooling configurations, we investigate how the temporal scale of the pooling data affects the quality of the forecasted precipitation.

Temperatures were not corrected as this study attempts to isolate the effects of post-processing precipitation, and correcting temperatures would then make the effects intertwined and difficult to evaluate separately. It was also shown previously that post-processing precipitation has a greater impact than post-processing temperatures (Verkade et al., 2013). The second case (forecasting with post-processed precipitation) was repeated three times: once for the Monthly QM, once for the Seasonal QM and once for the Annual QM.

2.7. Hydrological forecasting

The eight hydrological models were run for each day of the years 2015–2020 by using the ERA5 reanalysis data as the observed weather. For each day, optimal states of the eight hydrological models were identified, including the CemaNeige model states. The implementation of the EnKF over the period 2015–2020 leads to a distribution of 25 possible initial model states. Instead of preserving all 25 initial states, the mean value of the initial states was selected as the best estimate of actual initial states. This was done to remove the effects of the uncertainty related to the data assimilation and isolate the impacts of the ensemble weather forecasts on the hydrological forecast distribution. This representative state becomes the basis for applying post-processing techniques consistently, allowing for a more controlled and consistent analysis of the impact of post-processing on precipitation. The daily ECMWF weather forecasts were then used as inputs into the hydrological models, initializing the models with the assimilated initial states. The forecasted data included two scenarios: (1) the raw precipitation and temperature forecasts (i.e., as a control prior to the QM post-processing) and (2) the post-processed precipitation forecasts combined with the raw temperature forecasts. Using this methodology, a 50-member ensemble of streamflow forecasts is generated for a period of 10 days at the 6-hour time-step.

2.8. Performance evaluation

As suggested in multiple studies (e.g. McInerney et al. 2022), multiple metrics should be used to evaluate ensemble forecast performance. The Continuous Ranked Probability Score (CRPS) is one of the most well-known probabilistic scoring metrics (Matheson and Winkler, 1976) widely used in weather and hydrologic sciences. CRPS is used to assess the overall accuracy of competing forecasting systems by evaluating the distance between two distributions as:

$$CRPS = \int_{-\infty}^{\infty} [P(x) - P_0(x)]^2 dx \tag{4}$$

$$P_0(x) = \begin{cases} 0, x < x_0 \\ 1, x \geq x_0 \end{cases} \tag{5}$$

$P(x) = p_i \equiv \frac{i}{K}$, for $x_i < x < x_{i+1}$ (6)

where $\{x_1, \cdots, x_K\}$ are the ordered members of the ensemble of size K and x_0 is the observation.

In this study, we utilize the Talagrand diagram (rank histogram) as a valuable tool to visually depict the dispersion between forecasted runoff and observed data (Hamill, 2001). The Talagrand diagram helps assessing the dispersion of simulated runoff. A U-shaped diagram indicates under-dispersed simulated runoff, suggesting insufficient width to encompass observations. Consequently, observations tend to cluster at extreme percentiles, either the first or last, resulting in the characteristic U-shape. Conversely, a ∩ -shaped diagram indicates over-dispersion, implying excessive width. The presence of an “L” shape in the diagram signifies a systematic bias. These shapes, positioned mostly in the leftmost or rightmost bins of the histogram, occur when the forecast fails to adequately represent observations. The optimal diagram shape is flat, indicating a uniform distribution of observations within the forecasts. This configuration ensures an accurate representation of variability without introducing biases (Hamill, 2001).

The Average Bin Distance to Uniformity (ABDU) score was also added to quantitatively measure the deviation from uniformity, which is used in Arsenault et al. (2016) as the following Eq. (7):

$$\mathrm{ABDU} = \frac{1}{\mathrm{N}} \sum_{k=1}^{\mathrm{N}} \left| s_k - \frac{\mathrm{M}}{\mathrm{N}} \right| \tag{7}$$

where s_k is the number of occurrences in bin k, M is the number of y values, and N is the number of histogram bins. The term M/N is the expected value if the histogram is uniform and its value is fixed at 218.2 in this study (2182 data points in 10 bins). A larger ABDU reflects a larger deviation from uniformity. The ABDU is an empirical measurement to quantify the evaluation of the rank histogram and does not follow any particular distribution.

The last metric that was incorporated is the Normalized Mean Interquartile Range (NMIQR) index, defined in Bourgin et al. (2014). This metric helps to evaluate the sharpness of probabilistic forecasts. The interquartile range, defined as the range between the upper quartile (75th percentile) and the lower quartile (25th percentile) of a distribution, is a robust measure of the spread of a distribution. The Mean Interquartile Range (MIQR) is computed as:

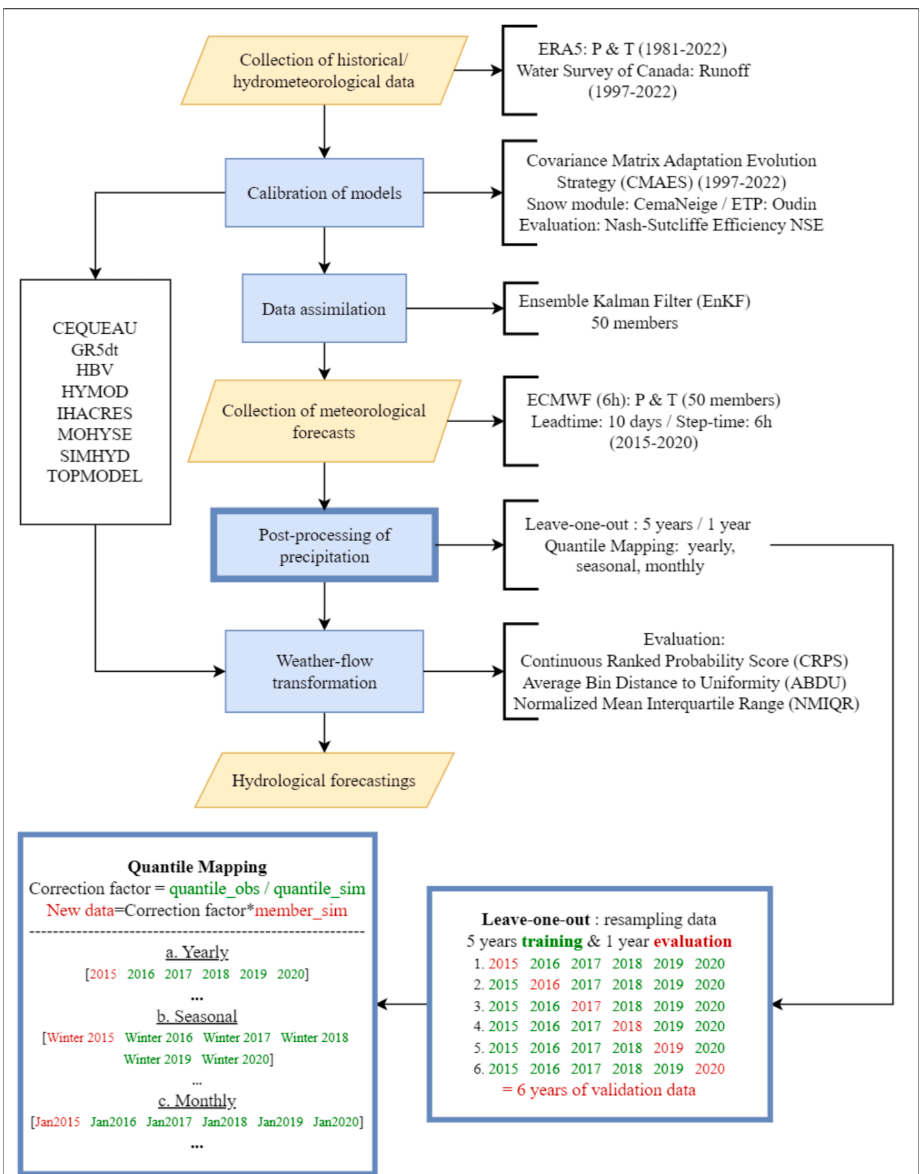

Fig. 2. Methodological overview, featuring visual representations of the QM and Leave-one-out techniques.

$$MIQR = \frac{1}{N}\sum_{k=1}^{N}\left(Q_{fct}^{75}(k) - Q_{fct}^{25}(k)\right)$$

where $(Q_{fct}^{25}(k), Q_{fct}^{75}(k))$ is the k^{th} of N pairs of quartiles of the forecasts. The MIQR is then divided by the mean runoff to obtain a non-dimensional score, becoming NMIQR. The closer the NMIQR index is to 0, the sharper the forecast. This metric was used to evaluate the sharpness performance of the forecasts and to understand the varying impact of the post-processing method on the ensembles.

All these steps are summarized in Fig. 2 to provide an overview of the methodological steps used in this study.

3. Results

In the following section, we present a selection of results that represent performance of the hydrological models used, the variability of the hydrological forecasting according to the post- processing method, lead-time and the ensemble of catchments.

3.1. Hydrological model performance on the sub-daily time-step

Firstly, Table 3 shows the Nash-Sutcliffe efficiency (NSE) metric of the performance of eight hydrological models with the twenty catchments over the calibration period.

Table 3 presents the NSE values obtained during the 1997–2022 period by hydrological model and per catchment. Most hydrological models provide good NSE values over the catchments (NSE>0.70). The best NSE values are obtained for Catchment 18. Low NSE values are obtained for catchment 4, where two models, GR5dt and TOPMODEL, displays poor performance at simulating streamflow, even after multiple calibration runs to ensure proper convergence was attained.

3.2. Evaluation of the post-processing methods

The performance of the three post-processing QM methods is evaluated by their ability to improve raw ECMWF precipitation forecasts over the 20 catchments for the 2015–2020 period. The Continuous Ranked Probability Score (CRPS) values were generated by (a) comparing the observed precipitation data with the forecasted precipitation data at each specific time step, and (b) comparing the observed cumulated precipitation data with the forecasted cumulated precipitation data at each specific time step, considering all forecast members for each time step individually. For the case (a), the precipitation forecasts were analysed at each 6-hour interval independently, without cumulating precipitation amounts over the forecast period. For example, a 24-hour lead time refers to forecasts made for the period from 18 to 24 h ahead. For the case (b) the precipitation forecasts were analysed at the accumulated precipitation over the forecast period. For example, a 24-hour lead time refers to forecasts made for the period from 0 to 24 h ahead. Figs. 3 to 5 present the precipitation CRPS values of 24 h, 4 days and 8 days of lead time over 2182 forecasts for each catchment (365 days per year, 366 days for 2016, and excluding the final 9 days of 2020 where no observed data is available in the following year to compare the forecasts).

We can see that the initial CRPS values of raw precipitations (see Fig. 3) are already low, below 0.2 mm per 6 h period on average. As the lead time increases, the median CRPS of raw precipitation values increases (see Fig. 5). In case (a) for all catchments, the three QM methods accomplish a substantial improvement compared to the raw ECMWF precipitation with a reduction of the median of the distribution of the mean CRPS values. In case (b), only a certain number of catchments are observed to accomplish a substantial improvement with the QM methods compared to the raw ECMWF precipitation with a reduction of the median of the distribution of the mean CRPS values.

From Supplementary Materials S1 – S4, a decrease in the inter-quartile range is obtained between the raw and post-processed precipitation forecasts for most catchments when looking at longer lead-times. The median of the distribution of the raw precipitation increases with the lead-time. The median of the distribution of the post-processed precipitation in Quebec watersheds (3 to 13) is higher than the median of raw precipitation from the first day of the lead time (24 h) and even the first 6 h. But for some watersheds, the median of the distribution of the post-processed precipitation is lower than the median of raw precipitation or it shows the same median values. Cumulative precipitation, however, shows that the impact of QM is less clear in some catchments and at longer lead-times. This points to the limits of QM for correcting precipitation amounts and to the ability to improve forecasts for shorter lead-times, on a timestep-by-timestep basis.

3.3. Effects of the post-processing on streamflow simulations

The three configurations of post-processed precipitations in addition to the raw forecasts were used as inputs to the hydrological models. The best performing hydrological model, based on the lowest median CRPS assessed by catchment, is presented in Table 4 and Figures S5 and S6 from Supplementary Materials.

Table 3
Nash-Sutcliffe efficiency (NSE) results during model calibration.

	Drainage area (km^2)	CEQUEAU	GR5dt	HBV	HYMOD	IHACRES	MOHYSE	SIMHYD	TOP-MODEL	Mean NSE
1	35,800	0.40	0.51	0.66	0.62	0.51	0.44	0.60	0.66	0.55
2	1030	0.71	0.74	0.71	0.69	0.70	0.65	0.63	0.67	0.69
3	1331	0.75	0.80	0.83	0.79	0.76	0.81	0.80	0.38	0.74
4	4560	0.66	0.19	0.63	0.67	0.63	0.59	0.72	**0.02**	**0.51**
5	5479	0.85	0.82	0.85	0.85	0.78	0.84	0.87	0.83	0.84
6	1163	0.82	0.80	0.82	0.83	0.80	0.84	0.84	0.81	0.82
7	1286	0.82	0.83	0.83	0.84	0.81	0.84	0.84	0.83	0.83
8	2163	0.80	0.81	0.83	0.82	0.79	0.80	0.82	0.80	0.81
9	708	0.73	0.77	0.78	0.78	0.74	0.76	0.76	0.76	0.76
10	5820	0.78	0.79	0.81	0.81	0.78	0.80	0.80	0.79	0.80
11	1152	0.78	0.76	0.78	0.78	0.75	0.78	0.78	0.77	0.77
12	821	0.73	0.42	0.75	0.76	0.73	0.76	0.76	0.77	0.71
13	696	0.69	0.70	0.72	0.72	0.69	0.71	0.72	0.71	0.71
14	1250	0.78	0.83	0.80	0.80	0.76	0.78	0.78	0.78	0.79
15	46.8	0.68	0.68	0.65	0.66	0.68	0.66	0.66	0.66	0.67
16	1350	0.74	0.78	0.75	0.73	0.74	0.63	0.74	0.61	0.71
17	780	0.86	0.81	0.84	0.87	0.84	0.80	0.87	0.80	0.84
18	4540	0.87	0.93	0.91	0.92	0.91	**0.94**	0.92	0.89	**0.91**
19	2310	0.73	0.82	0.76	0.76	0.84	0.84	0.81	0.79	0.79
20	33.6	0.61	0.66	0.62	0.61	0.64	0.61	0.60	0.56	0.62
Mean NSE		**0.74**	**0.72**	**0.77**	**0.76**	**0.74**	**0.74**	**0.76**	**0.70**	

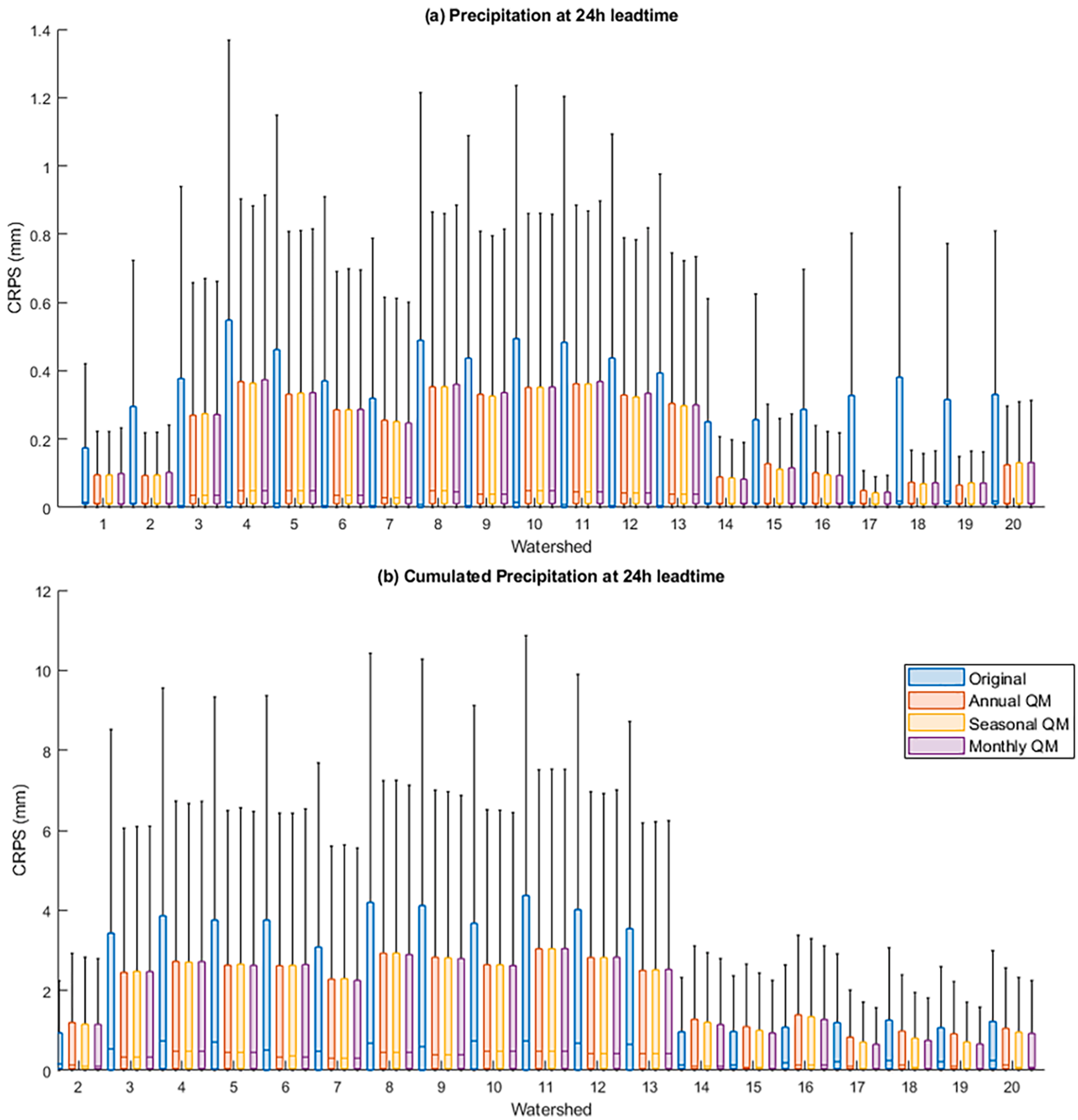

Fig. 3. Boxplots of the (a) precipitation and (b) cumulated precipitation CRPS of the 20 catchments used in this study at the 24 h lead time for the four precipitation sets used in this study: Raw ECMWF precipitation (blue), Annual QM (orange), Monthly QM (yellow) and Seasonal QM (purple). Each boxplot (using the 5th, 25th, 50th, 75th, and 95th percentiles) summarizes the variety of CRPS scores over all 2182 forecasts at a 24-hour lead time.

GR5dt and TOPMODEL seem to provide best performance for the southeastern Canadian catchments (Figure S5). In contrast, the rest of hydrological models (CEQUEAU, HBV, HYMOD, IHACRES, MOHYSE, SIMHYD) seem not to have specific abilities for a particular Canadian region. The choice of the hydrological model also seems to not depend on precipitation input (raw ECMWF, annual QM, monthly QM or seasonal QM) for most catchments. Table 4 shows the most effective hydrological model at each catchment by post-processing implementation.

In Table 4, the best hydrological model for each watershed is indicated based on its median CRPS value. The hydrological model with the lowest median CRPS value is listed in the table. Below each hydrological model name, a percentage is provided, indicating how much lower the median CRPS value is compared to the other hydrological models. For instance, in catchment 20, the TOPMODEL is identified as the optimal choice for raw precipitation, exhibiting a median CRPS value that was approximately 24.4 % lower, on average, than the median CRPS values of the other hydrological models. For some catchments, the optimal hydrological model for a given catchment is typically narrowed down to two models (see catchments 11, 14, 15, 17 and 19 from Table 4), based on their performance when subjected to the varying precipitation inputs.

The pie graph in Figure S6) shows the fraction of all cases for which each hydrological model performs best. From these results, Table 4 and S5 and S6, it is clear that the hydrological model GR5dt is the preferred choice for most catchments (>50 %). In contrast, the hydrological model TOPMODEL exhibits an intermediate preference, with an average favourability rating of approximately 30 %. Intriguingly, HYMOD and IHACRES fail to secure the top position for any watershed or precipitation configuration.

From the CRPS results, three distinct groups of results among the 20 watersheds analysed are identified. From each of these groups, one watershed is aleatory selected. Therefore, a further analysis of the mean CRPS values per lead-time is provided for three representative catchments: 40,624 – Quebec (Fig. 6); 03OE011 – Newfoundland and Labrador (Fig. 7), and 01EF001 – Nova Scotia (Fig. 8). The results for the remaining catchments can be found in Supplementary Materials S7-S23.

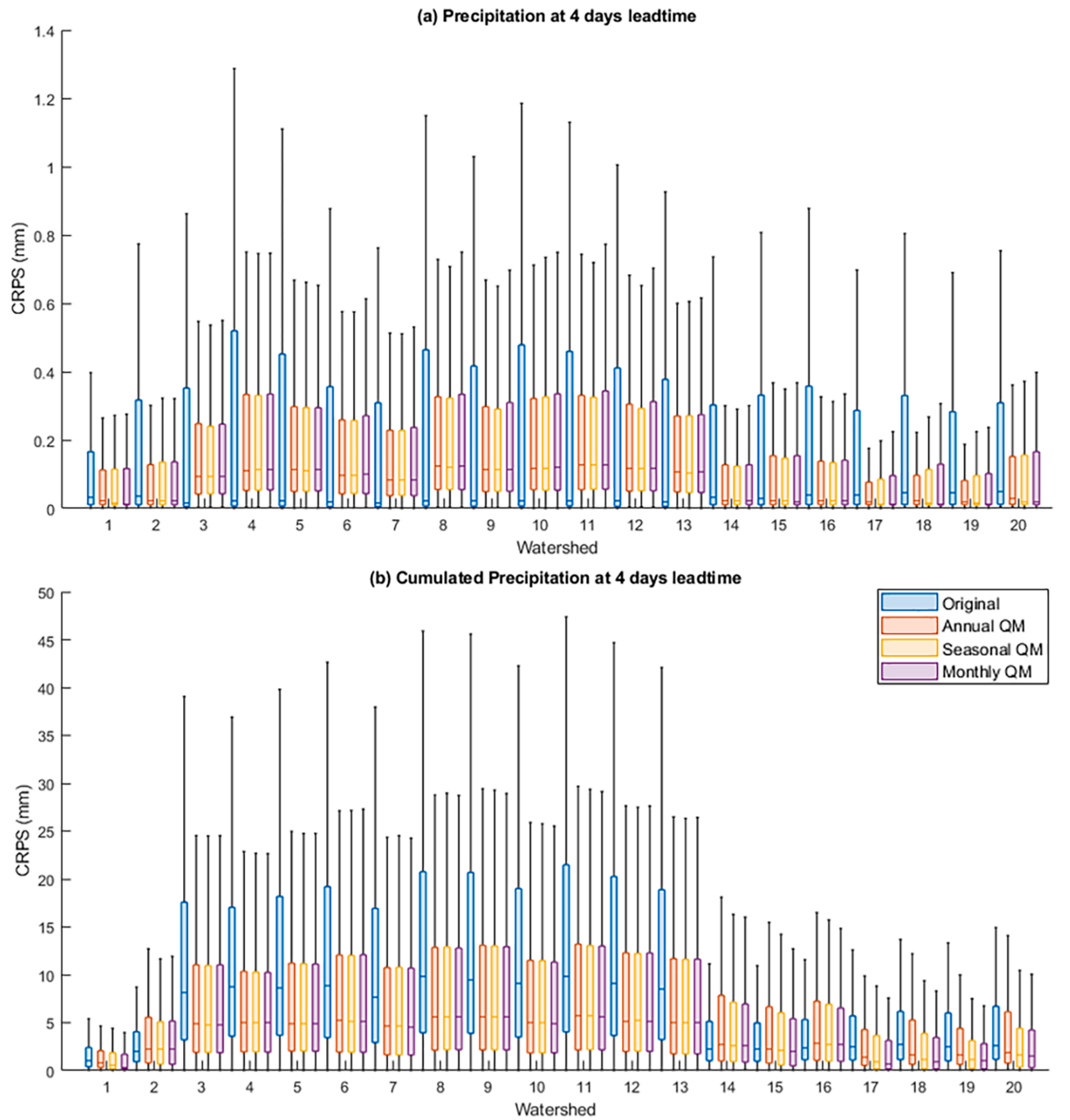

Fig. 4. Boxplots of the (a) precipitation and (b) cumulated precipitation CRPS of the 20 catchments used in this study at the 4 days lead time for the four precipitation sets used in this study: Raw ECMWF precipitation (blue), Annual QM (orange), Monthly QM (yellow) and Seasonal QM (purple Each boxplot (using the 5th, 25th, 50th, 75th, and 95th percentiles) summarizes the variety of CRPS scores over all 2182 forecasts at a 4 days lead time.

From Fig. 6 it is possible to observe that the application of the annual QM post-processing method leads to improvements in terms of bias and skill in the forecasted streamflow ensembles, particularly at longer lead times and for a selected group hydrological models in catchment 4. This behaviour is observed for 11 out of the 20 stations (55 %) (refer to Supplementary Materials for similar figures for all catchments). It is also observable that the performance of the QM implementation (annual, seasonal or monthly) varies according to the hydrological model used. These catchments can be identified with a green colour in Table 4.

Fig. 7 shows that the application of the post-processing techniques shows no effect on most of the CRPS median values at the catchment 17 – station 03OE011. Some increases and some decreases in the CRPS scores are seen for the three QM configurations at longer lead-times on some hydrological models. This behaviour is observed for 7 out of the 20 stations (35 %), as seen in the supplementary materials. These catchments can be identified with a yellow colour in Table 4.

Finally, Fig. 8 shows an increase of the median and spread of CRPS values for the catchment 14 – 01EF001 as the lead time increase for the three QM configurations and all the hydrological models. This behaviour is observed in 2 out of the 20 stations (10 %). These catchments can be identified with a red colour in Table 4.

These findings suggest that the QM configuration has a positive impact on the forecast accuracy and reliability in a half of the catchments (55 %). It is also observable that for 35 % of watersheds, depending on the hydrological model used, the QM (annual, monthly, or seasonal) can slightly decrease, increase or have no change on the CRPS of the forecasts. And finally, the use of QM provides an increase of the CRPS in 10 % of catchments used in this study (See Supplementary Materials S7-S23 and Table 4).

Subsequently, Figs. 9 and 10 display the Talagrand diagrams for watershed 40,624 (catchment 4). These diagrams exclusively feature the hydrological models GR5dt and SIMHYD, and encompass four distinct lead-times: 24 h, 2 days, 4 days, and 8 days. The U-shaped configuration of the diagrams denotes under-dispersion, and interestingly, as lead-

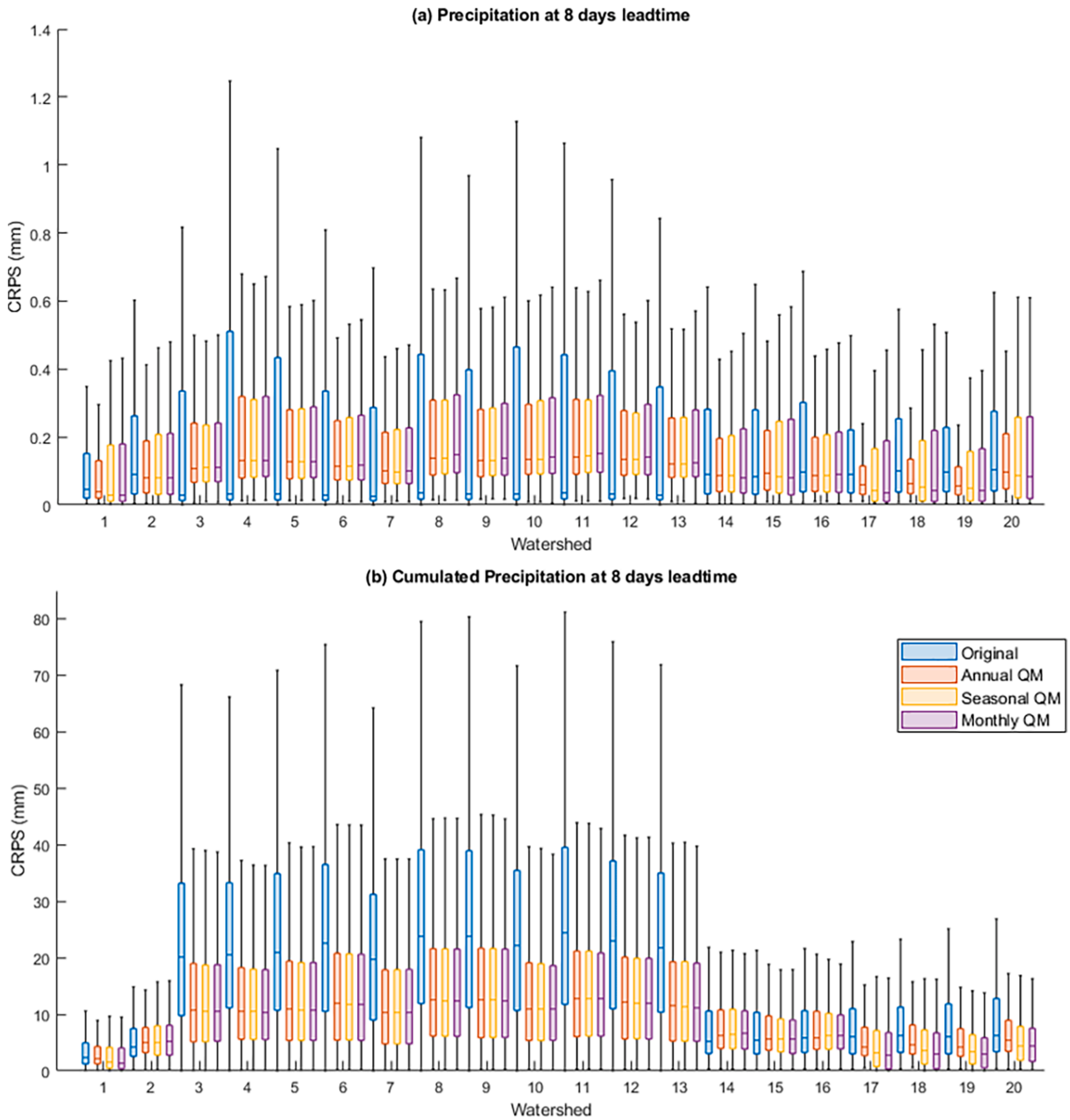

Fig. 5. Boxplots of the (a) precipitation and (b) cumulated precipitation CRPS of the 20 catchments used in this study at the 8 days lead time for the four precipitation sets used in this study: Raw ECMWF precipitation (blue), Annual QM (orange), Monthly QM (yellow) and Seasonal QM (purple). Each boxplot (using the 5th, 25th, 50th, 75th, and 95th percentiles) summarizes the variety of CRPS scores over all 2182 forecasts at a 8 days lead time.

time increases, there is a subtle improvement in shape when applying Quantile Mapping.

Recalling the ideal ABDU value for this study is 218.2, the results from GR5dt showcase an enhancement in the ABDU value with increasing lead-time. Notably, for this specific watershed, SIMHYD emerges as the most proficient hydrological model. It is noteworthy that the ABDU values associated with SIMHYD consistently exhibit lower values compared to those obtained using GR5dt, emphasizing the superior performance of SIMHYD in this particular watershed.

Finally, Fig. 11 display the distributions of the sharpness index, the normalized mean interquartile range (NMIQR), of the ensemble of 8 hydrological models and the 20 catchments.

Recalling that the closer the NMIQR index is to 0, the sharper the forecast, it is easy to appreciate that the raw forecasts have a very low NMIQR value. As the lead time increases, the sharpness decreases. Forecasts that used a QM precipitation input tend to be less sharp than the raw forecasts, especially from the 4th day of the forecast.

4. Discussion

4.1. Hydrological model performance and sensitivity to uncertainty

The eight hydrological models were applied at a 6-hourly timestep to better simulate streamflow on the sub-daily timesteps. With few exceptions, all the hydrological models were calibrated with at least satisfactory results, with NSE values ranging from 0.51 to 0.91 in calibration, and most models were able to perform relatively well (with NSE values > 0.7) on most catchments. Some exceptions exist for a subset of catchment-model combinations, but these are limited overall. Forecasts for up to 9 days ahead are provided at the 6-hourly time-step. A sub-daily timestep helps better represent certain hydrological processes such as snowmelt and evapotranspiration, which are highly impacted by the diurnal cycle. This allows capturing the physical processes more precisely than at the daily time step while improving the models' ability to forecast streamflow on the study catchments. For example, snowmelt

Table 4
Most effective hydrological model at each catchment by post-processing method.

	Province	Drainage area (km^2)	CRPS Best Hydrological Model				QM Performance
			Raw ECMWF	Annual QM	Monthly QM	Seasonal QM	
1	AB	35,800	HBV ~ 48.1 %	HBV ~ 45.2 %	HBV ~ 45.2 %	HBV ~ 46.1 %	
2	ON	1030	GR5dt ~ 25.9 %	GR5dt ~ 25.6 %	GR5dt ~ 24.3 %	GR5dt ~ 23.5 %	
3	QC	1331	GR5dt ~ 30.0 %	GR5dt ~ 29.7 %	GR5dt ~ 29.3 %	GR5dt ~ 29.3 %	
4	QC	4560	SIMHYD ~ 35.0 %	SIMHYD ~ 30.3 %	SIMHYD ~ 30.4 %	SIMHYD ~ 30.7 %	
5	QC	5479	GR5dt ~ 17.1 %	GR5dt ~ 19.1 %	GR5dt ~ 19.2 %	GR5dt ~ 18.5 %	
6	QC	1163	GR5dt ~ 18.8 %	GR5dt ~ 15.5 %	GR5dt ~ 15.8 %	GR5dt ~ 15.8 %	
7	QC	1286	GR5dt ~ 25.3 %	GR5dt ~ 21.4 %	GR5dt ~ 21.7 %	GR5dt ~ 21.7 %	
8	QC	2163	GR5dt ~ 21.7 %	GR5dt ~ 16.7 %	GR5dt ~ 17.1 %	GR5dt ~ 17.4 %	
9	QC	708	TOPMODEL ~ 30.3 %	TOPMODEL ~ 14.3 %	TOPMODEL ~ 14.3 %	TOPMODEL ~ 15.4 %	
10	QC	5820	GR5dt ~ 28.7 %	GR5dt ~ 19.5 %	GR5dt ~ 19.4 %	GR5dt ~ 19.7 %	
11	QC	1152	TOPMODEL ~ 25.3 %	GR5dt ~ 12.8 %	GR5dt ~ 13.3 %	GR5dt ~ 12.1 %	
12	QC	821	TOPMODEL ~ 31.4 %	TOPMODEL ~ 19.5 %	TOPMODEL ~ 19.4 %	TOPMODEL ~ 19.9 %	
13	QC	696	GR5dt ~ 14.7 %	GR5dt ~ 6.7 %	GR5dt ~ 6.5 %	GR5dt ~ 6.4 %	
14	NS	1250	TOPMODEL ~ 14.9 %	GR5dt ~ 14.5 %	TOPMODEL ~ 12.4 %	TOPMODEL ~ 11.8 %	
15	PE	46.8	TOPMODEL ~ 16.6 %	GR5dt ~ 17.1 %	TOPMODEL ~ 16.4 %	TOPMODEL ~ 18.7 %	
16	NS	1350	GR5dt ~ 22.7 %	GR5dt ~ 29.8 %	GR5dt ~ 28.2 %	GR5dt ~ 26.3 %	
17	NL	780	GR5dt ~ 16.4 %	GR5dt ~ 18.7 %	CEQUEAU ~ 17.9 %	CEQUEAU ~ 15.4 %	
18	NL	4540	TOPMODEL ~ 36.8 %	TOPMODEL ~ 34.3 %	TOPMODEL ~ 34.8 %	TOPMODEL ~ 35.0 %	
19	NL	2310	MOHYSE ~ 17.4 %	MOHYSE ~ 22.3 %	MOHYSE ~ 14.6 %	TOPMODEL ~ 17.5 %	
20	NL	33.6	TOPMODEL ~ 24.4 %	TOPMODEL ~ 21.9 %	TOPMODEL ~ 27.3 %	TOPMODEL ~ 26.3 %	
	Model		**GR5dt (10/20)**	**GR5dt (13/20)**	**GR5dt (10/20)**	**GR5dt (10/20)**	

was computed 4 times per day, leading to a more realistic snowmelt rate, with higher snowmelt values during the daytime than at night.

We showed that certain hydrological models might exhibit superior performance in simulating sub-daily streamflow compared to others. Specifically, the GR5dt hydrological model emerges as the top performer in approximately half of the catchments, accounting for approximately 55 % of cases. However, it's worth noting that while these models may exhibit the smallest median CRPS values, they typically demonstrate only a 10 to 30 % reduction compared to alternative hydrological models (see Table 4). This suggests that other models also perform commendably, a conclusion supported by the NSE values derived from calibration efforts.

The set of hydrological models used in this study has a similar model structure (i.e., lumped models with 5 to 9 parameters based on the same PET and snow module routines), however some differences can be observed between models for some catchments (Table 3 and Table 4). Notably, GR5dt and TOPMODEL, featuring 5 and 7 parameters respectively, generally emerge as the top two performing models, but with ~ 10 to 30 % of smaller CRPS medians compared to the rest. On the other hand, other hydrological models utilizing 9 and 8 parameters (HBV and SIMHYD), when selected once each as best performers, exhibit >~ 30—40 % of lower CRPS medians. Therefore, the apparent skill of the various hydrological models is not found to be related to their complexity, as simpler models (such as GR5dt with 5 parameters and 3 reservoirs) often outperform models such as HBV (9 parameters and 3 reservoirs), and vice-versa. For the model calibration, based on the NSE scores (Table 3), the most-likely best performing hydrological model is TOPMODEL for catchment number 12, however this model leads to unsatisfactory NSE results for catchment number 4. To ensure that this was not due to the calibration convergence, calibration was performed 10 times for each model and the parameter set with the best NSE value was preserved. This suggests that the differences between models are mostly caused by their internal hydrological process representation, which agrees with Thiboult and Anctil (2015), who found that the model structure and conceptualization are the dominant sources of uncertainty.

When we look at S5 (Supplementary Materials), we can see that the distribution of best models shows a pattern of preference: GR5dt for the

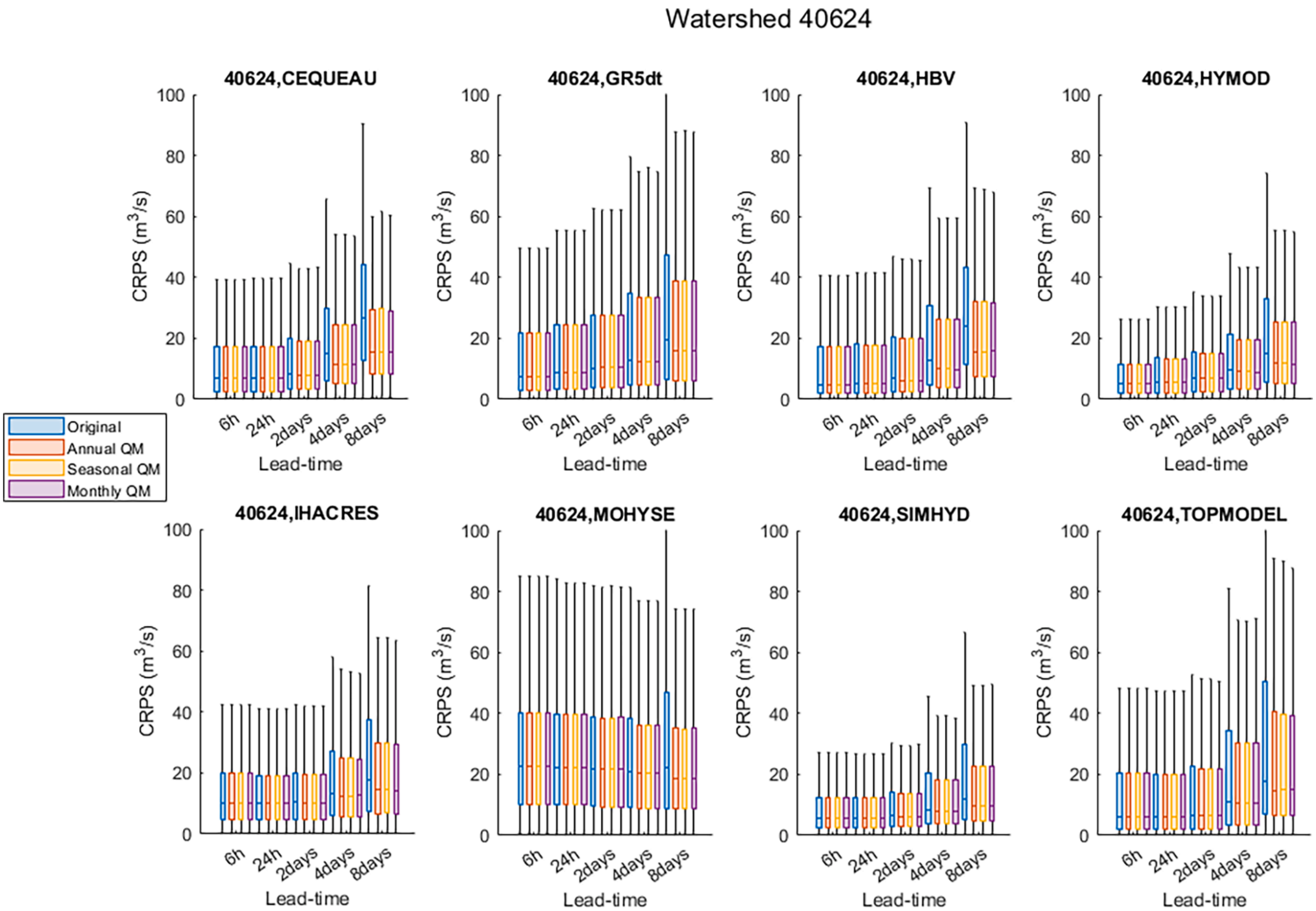

Fig. 6. CRPS values of the 2015–2020 hydrological forecasting per lead-time for the 40,624 station (catchment 4) with the eight different hydrological models. Each boxplot (using the 5th, 25th, 50th, 75th, and 95th percentiles) summarizes the variety of CRPS scores over all 2182 forecasts.

southern Canada catchments and TOPMODEL for the northeaster Canada. From the catchments that are located in the province of Quebec, all prefer GR5dt and TOPMODEL except for the 4th catchment that prefers SIMHYD. Interestingly, this catchment is the northernmost of that province. This could be explained by the sizes of the different catchments and the hydrologic regime. The catchments in eastern Canada, are small to medium-size catchments (33——5820 km^2) while the only large-scale catchment is located in the western part (35800 km^2). It could also be explained by the region, topology or land use of the different catchments. The latter are complex to model in western catchments due to the flat lands that make the effective drainage areas variable according to the precipitation intensity (see Shaw et al., 2013; Muhammad et al., 2019), leading to high CRPS values. In addition, flowrates in the prairie regions are very small, with most water being evaporated before reaching the water course.

The diverse performances observed among the hydrological models raise important questions regarding the underlying factors driving these differences and why certain models outperform others. Despite dedicated analysis, no specific model characteristics or attributes have been identified as consistently correlating with performance outcomes. This observation underscores the complexity of hydrological modeling and highlights the need for robust strategies to account for model uncertainty. As suggested by Troin et al. (2021), the use of model ensembles may offer a promising approach to address these challenges by capturing a broader range of model behaviors and improving the reliability of hydrological forecasts.

4.2. Post-processing of precipitation forecasts and process-conditioning

The post-processing of precipitation forecasts was implemented to attempt to improve streamflow forecasts. The three QM implementations improve precipitation forecasts accuracy compared to the raw forecasts in most catchments, with a decrease in the CRPS score (Figs. 3 to 5). While QM demonstrates a positive impact on precipitation forecasts across all watersheds, the assessment of cumulative precipitation reveals a more nuanced picture, with QM showing a positive effect in only half of the watersheds. The consideration of CRPS in precipitation and cumulative precipitation enables us to not only evaluate the immediate impact of QM on individual precipitation events but also to observe the accumulated error over time. This distinction highlights that while QM may effectively correct certain aspects of precipitation forecasts in some watersheds, its efficacy in improving cumulative precipitation and subsequent hydrological responses varies across different watershed contexts. By examining cumulative precipitation, we gain deeper insights into the overall performance and limitations of QM across diverse hydrological settings.

As the QM is an unconditional post-processing method, the improvement shown in half of the watersheds can be attributed to the correction of long-term precipitation biases. These improvements are low even in the raw forecast (CRPS usually below 1 mm) in large part because the forecasted precipitation is generated using the ECMWF forecast model that is also used to generate the ERA5 reanalysis, which is used as observations. Therefore, it is expected that the forecasted precipitation should be of high quality for the first few lead-times, where the atmospheric assimilation is similar between both methods. This was

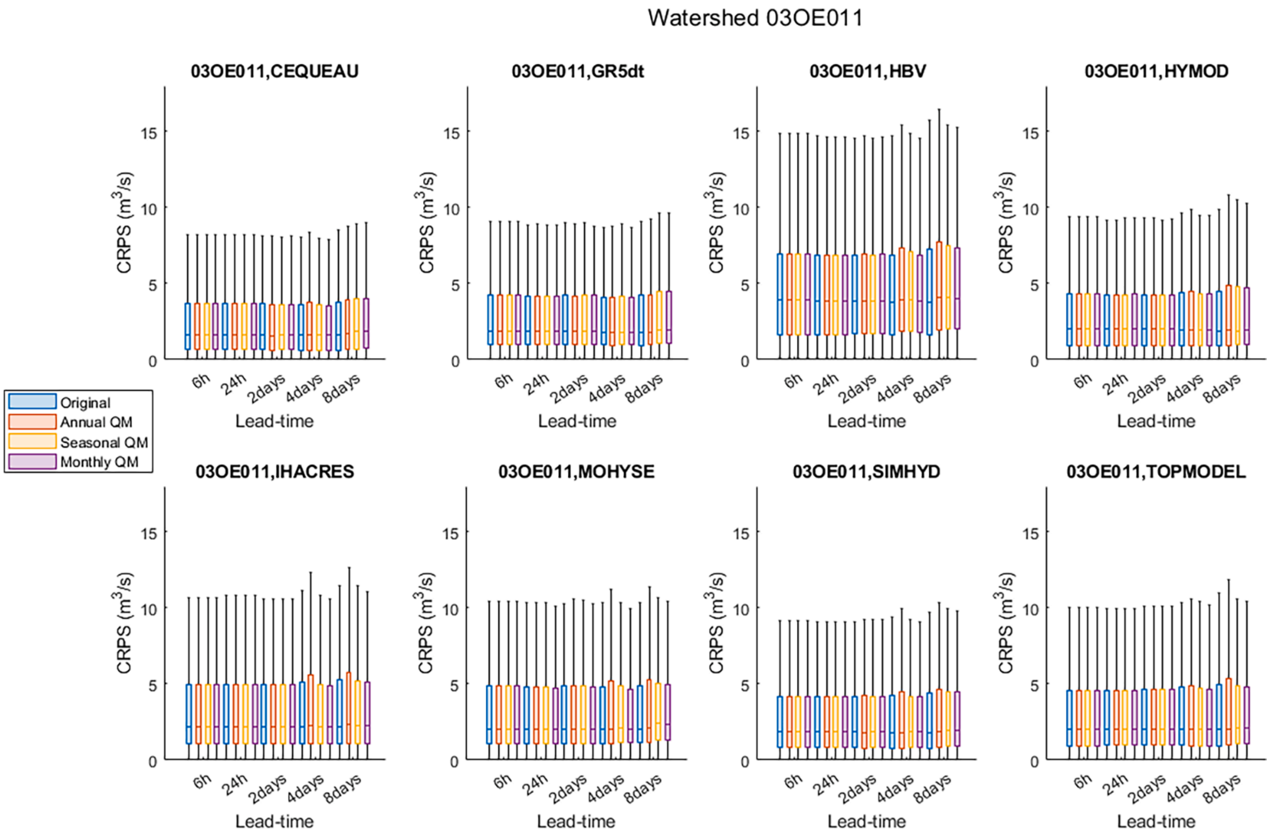

Fig. 7. CRPS values of the 2015–2020 hydrological forecasts per lead-time for the 03OE011 station (catchment 17) with the eight different hydrological models. Each boxplot (using the 5th, 25th, 50th, 75th, and 95th percentiles) summarizes the variety of CRPS scores over all 2182 forecasts.

indeed the case, as shown in figures S1 to S4 in the Supplementary Materials. Results might differ in cases where the observation and forecast models are less similar, due to the time gap of data (±2h) between the ECMWF forecasts (issued every 6 h UTC based) and the local observation times in certain basins.

In this study, we chose to use the precipitation as the conditioning variable (or proxy) for hydrological processes. In separating by season and by month, the QM parameters were computed on days within that same season / month, ensuring consistency in hydrometeorological conditions. We showed that conditioning on the sub-daily precipitation was able to reduce higher values of CRPS associated with raw precipitation, albeit at the expense of an increased median CRPS. This trend is particularly evident at longer lead times, emphasizing the trade-off between mitigating extreme values and maintaining overall forecast accuracy. However, when comparing the three QM conditioning configurations for sub-seasonal precipitation (annual, monthly, or seasonal), no single configuration exhibits a clear advantage in forecasting results across most catchment basins. The optimal QM configuration varies among basins, emphasizing the need for basin-specific considerations in choosing the appropriate configuration. While there is a subtle indication that the annual configuration may perform marginally better at longer lead times, the contrast is not pronounced.

An open question remains regarding how to choose the variables to condition and the temporality of such conditioning. While it is theoretically possible to include more variables to condition on (e.g., temperature as a proxy for snow processes), this becomes challenging due to the issue of dimensionality leading to estimated sparse and noisy PDFs (Bennett et al., 2022). In the specific case for Canada, the hydrology of catchments is heavily influenced by the accumulation and subsequent melting of snow during the winter and spring seasons. However, the current snow model, CemaNeige, is considered to be overly simplistic as it relies solely on temperature inputs (Troin et al., 2016). This simplicity limits its ability to capture the complexity and uncertainty associated with snow hydrology. To address this, there is a growing recognition of the need to explore more advanced snow models that can provide higher temporal and spatial resolution simulations of the snow water equivalent (SWE), combining the accuracy of physically-based energy-balance (EB) models with the simplicity of degree-day (DD) models, resulting in mixed EB/DD models (Troin et al., 2016). Therefore, future studies should investigate more complex conditional post-processing methods to attempt to improve more specific and conditional biases.

4.3. Impact on precipitation post-processing on streamflow simulations

In this study, we focused on the precipitation postprocessing and the associated impacts on streamflow forecasts; we did not apply QM directly on the streamflow forecasts as this would mask the effects of precipitation postprocessing (Lucatero et al., 2018). A similar study by Ghimire et al. (2019) used seven parametric and nonparametric QM techniques to improve hydrological simulations. The QM techniques were applied annually, monthly, and seasonally to precipitation as in the present study. Ghimire et al. (2019) showed that the non parametric empirical quantile mapping methods yielded a very good hydrological performance at all temporal scales when applied under the monthly approach. Based on the CRPS scores obtained in our study, we show that out of the three QM configurations used, none outperforms other configurations. This lack of significant differentiation among the configurations is also evident in the Talagrand diagrams of Figs. 9 and 10, where there is minimal variation in the distribution among the three QM configurations. However, the QM does show improvement of flow forecasts in the catchments located in Quebec with most hydrological models. The fact that the QM method improves results for most test-cases agrees with the results found in Ghimire et al. (2019). However, that application was made on a single catchment and a single

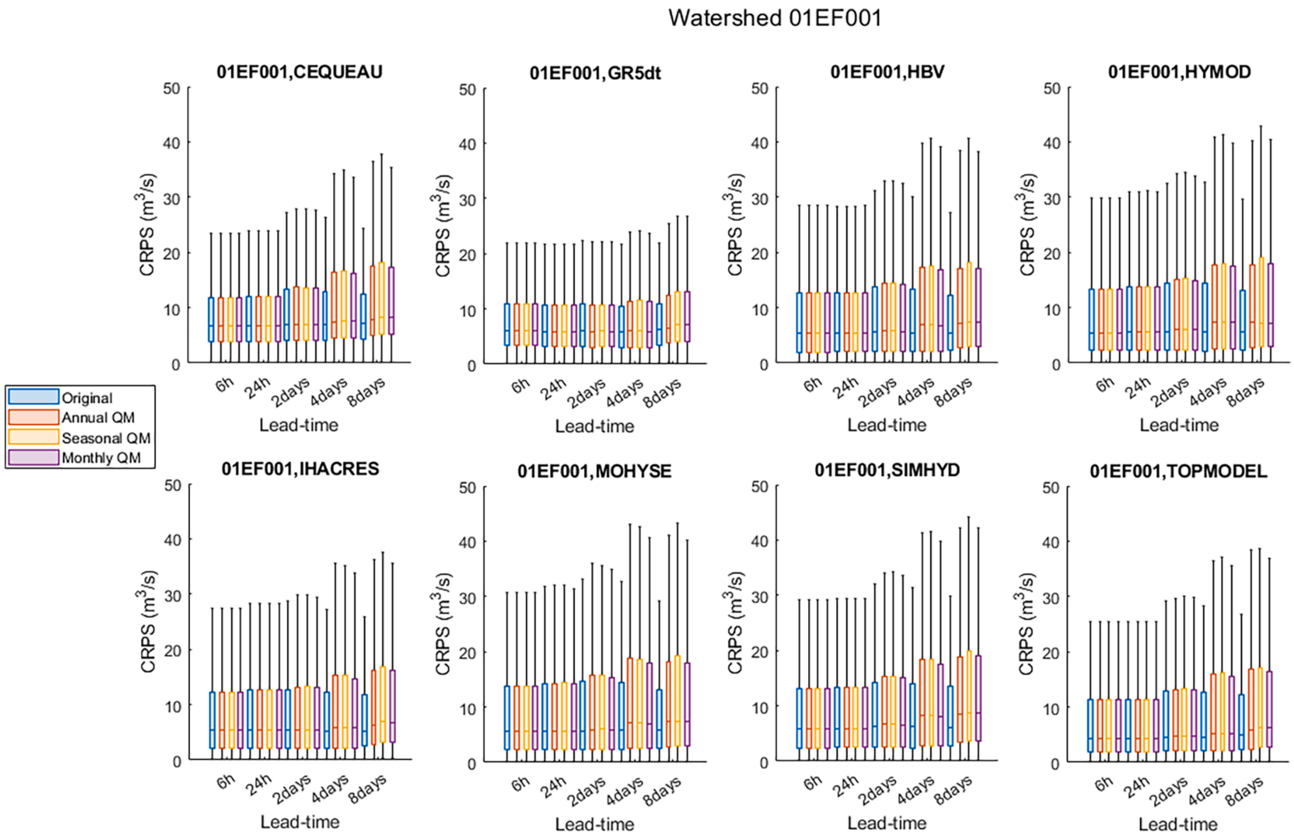

Fig. 8. CRPS values of the 2015–2020 hydrological forecasts per lead-time for the 01EF001 station (catchment 14) with the eight different hydrological models. Each boxplot (using the 5th, 25th, 50th, 75th, and 95th percentiles) summarizes the variety of CRPS scores over all 2182 forecasts.

hydrological model (GR4J), limiting the generalisation of the results.

Despite demonstrating satisfactory performance for half of the catchments, the hydrological models exhibit less favorable streamflow simulations for the remaining catchments when subjected to the three QM configurations. This disparity is particularly evident in Figs. 7 and 8, as well as Supplementary Materials S7, S8, and S19-S23. Surprisingly, there is no discernible difference on CRPS values, as it doesn't make any effect on the forecasted streamflow, and even an increase of CRPS is seen at longer lead times with some hydrological models. Also, the reduction in sharpness is another dimension to consider alongside overall performance improvements. As seen in Fig. 11, forecasts with QM precipitation tend to be less sharp than the raw forecasts, especially from the 4th day of the forecast, as indicated by higher NMIQR values.

Perhaps this observed behavior can be attributed to the hydrological regime of the watersheds. Figures S24 - S26 in Supplementary Materials illustrate the hydrological regimes of 6 different watersheds. S24 illustrates two out of eleven watersheds, all from the province of Quebec, that demonstrated a positive improvement with the implementation of QM. In S25, hydrographs from two representative watersheds out of the seven watersheds where QM had no discernible effect on streamflow forecasts. Finally, S26 displays hydrographs from the two watersheds that experienced a negative impact with the application of QM.

The hydrographs in S24 depict a pronounced peak in runoff occurring around 120 days, roughly corresponding to the month of April, which corresponds to the ablation of snow. Similarly, the hydrographs in S25 exhibit a similar pattern, albeit with higher peaks exceeding 800 m^3/s and minimal runoff at the onset of the year. In contrast, the hydrographs in S26 display a distinct runoff regime, featuring two significant periods of elevated runoff—one at the beginning and another towards the end of the year, marked by multiple peaks around this period.

The substantial peaks in runoff may contribute to the limited efficacy of QM in improving forecasts. When applying QM, timing errors tend to be averaged out, especially at longer lead times. In the short term (i.e., up to 1–2 days of lead time at 6-hour increments), even minor errors in peak flow timing can result in an increase in the CRPS, whereas these errors tend to offset each other over an extended period.

The outcomes of this study suggest that, despite its simplicity for operational forecasting streamflow purposes (Boucher et al., 2015), the implemented QM method proves ineffective as a universal post-processing technique for any catchment in this study. A prior analysis of catchment characteristics, including hydrological regime, topology, location, among others, is imperative to ascertain its viability as a suitable alternative.

4.4. Limitations and implications for hydrological studies

The experimental design of this study has some limitations. First, only twenty catchments are used for the analysis and additional catchments in different locations with different hydroclimatic characteristics in Canada should be included. As mentioned in section 2.1, the initial datasets that were made available were limited in spatial distribution and most hydrometric gauges with data at the hourly or sub-daily scales were regulated rivers, reducing the number of catchments used for this study. Therefore, other regions with more readily accessible datasets could replicate this study with a more diverse set of catchments.

Also, we have addressed the assumption of applying the same methodology to a large watershed as to the rest of the watersheds. It's important to recognize that hydrological processes in large watersheds, such as the one utilized in our study, may exhibit significantly slower responses compared to smaller catchments. The effect of different scales on inherent characteristics of the bigger catchment, such as residence times, drainage, soil depth and aquifer depth, etc., poses challenges when evaluating sub-daily forecasts, as the temporal resolution may not

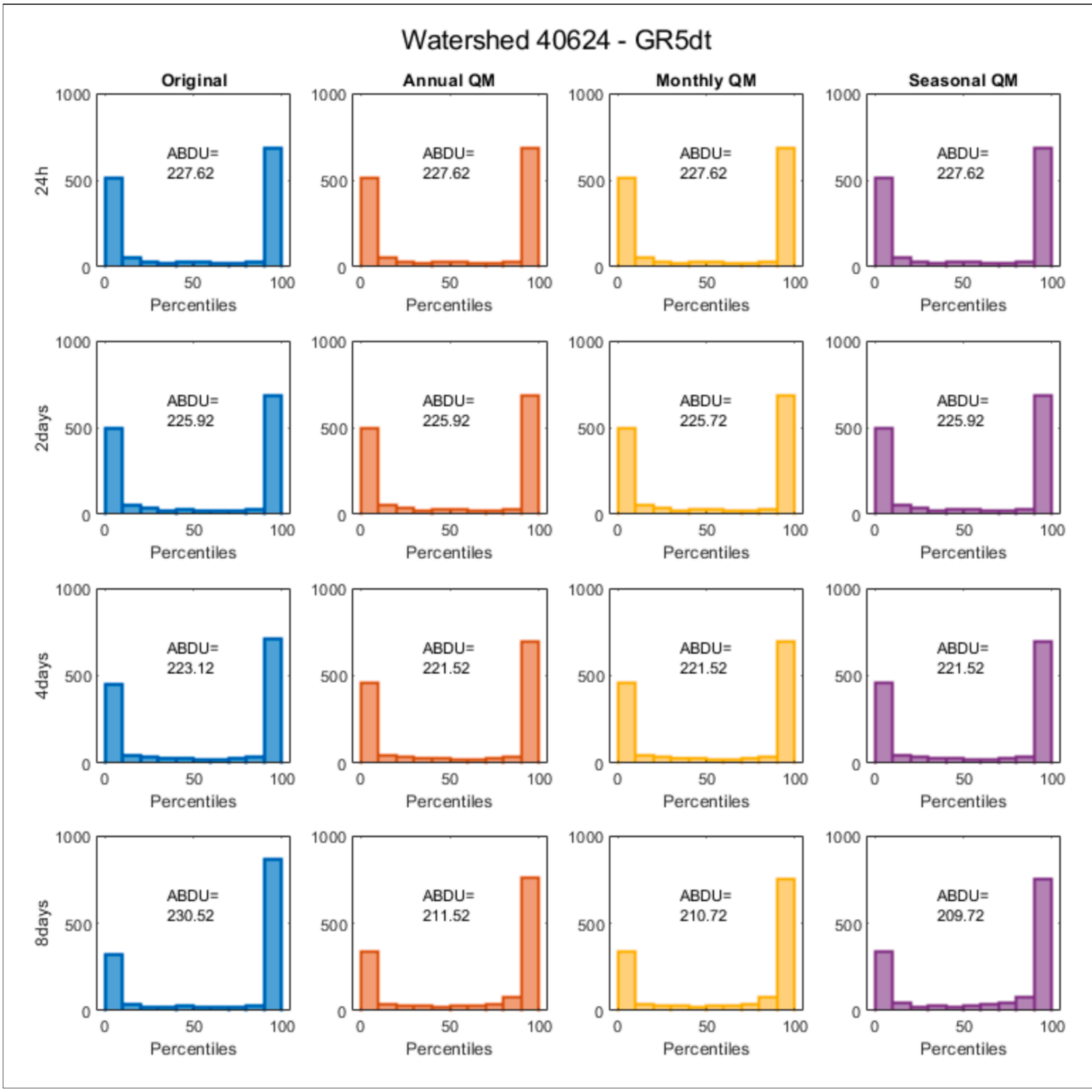

Fig. 9. Talagrands per lead-time of the GR5dt forecasts for the 40,624 station (catchment 4) with the 4 different QM configurations.

be adequate to capture subtle differences or responses.

Even though conceptual hydrological models are often used in streamflow forecasting studies (Arsenault et al., 2015; Dion et al., 2021; Velazquez et al., 2011; Andraos and Najem, 2020), the inclusion of one or more hydrological models in the ensemble with various representations of key hydrological processes (i.e., snow, potential evapotranspiration) would likely influence the results while allowing to extend the analysis of hydrological uncertainty associated with the model structure (Troin et al., 2016; Seiller and Anctil, 2016). Using distributed or physical-based hydrological models would provide a more thorough assessment of the impacts of precipitation post-processing on streamflow forecasts, especially for complex catchments such as in mountainous areas, for which the improvement of sub-daily precipitation forecasts is required.

As for the performance metric, using a single efficiency criterion, such as CRPS, may not be sufficient for evaluating the performance of hydrological models. Evaluating model performance involves subjective and objective assessments, and the choice of performance metrics can be challenging due to factors such as the variability of flows, heteroscedastic errors, different benchmark models, and specific model applications (Krause et al., 2005; Pushpalatha et al., 2012; Ferreira et al., 2020). While visual inspection of the CRPS boxplots provides qualitative insights, only one metric criterion may overestimate or underestimate different types of errors. Therefore, a combination of multiple criteria, as the use of more metrics such as KGE (Gupta et al., 2009), could help to analyse better the actual performance of the proposed methodology. However, in the context of hydrological forecasting, the choice of an objective function is not as critical as in general streamflow simulation, as the relationship between forecast performance and most objective functions used in calibration is tenuous at best and varies according to catchment, model, and other factors (Jie et al., 2016). Despite this, we chose to use NSE in calibration as it is still commonly employed in hydrological studies, facilitating comparisons with previous research, even though the metric itself is not used for direct evaluation purposes in this

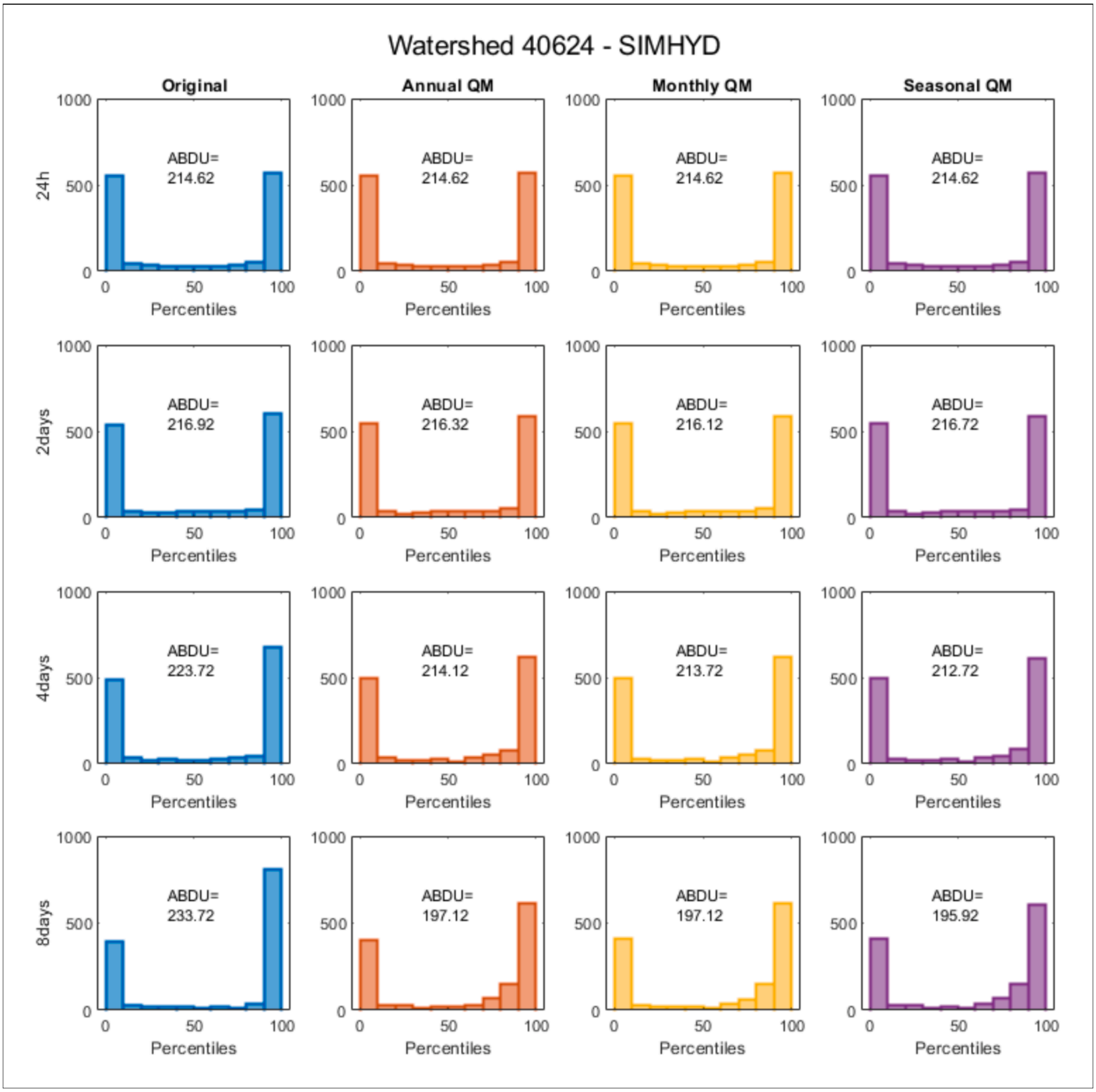

Fig. 10. Talagrands per lead-time of the SIMHYD forecasts for the 40,624 station (catchment 4) with the 4 different QM configurations.

study.

Also, while our study focused on the application of Quantile Mapping under stationary conditions (training period over the evaluation period), the potential impact of non-stationarity on the effectiveness of QM remains an important consideration. Addressing non-stationary conditions, such as those induced by climate change, represents a critical aspect that warrants further investigation (Chen et al., 2019). Future research could explore the robustness of QM techniques in accommodating changing climatic conditions and their implications for hydrological forecasting accuracy.

Finally, this study did not explore the application of statistical post-processing techniques on the forecasted streamflow to improve the accuracy of the hydrological forecast. Both weather and streamflow post-processing methods can potentially enhance the quality of streamflow forecasts. It would be valuable to assess the combined impact of these post-processing methods on the resulting hydrological forecasts. Previous studies have explored similar analyses but on smaller scales and using daily models, which are less complex compared to sub-daily models that capture more intricate hydrological processes (Li et al., 2018; Liu et al., 2022).

5. Conclusions and recommendations

This study aimed to assess the effectiveness of the application of Quantile Mapping (QM) as a post-processing method for precipitation forecasts to enhance hydrological forecasts. Three temporal configurations (monthly, seasonal, and annual) of the quantile mapping scheme were specifically employed. The investigation focused on 20 catchments located in Canada, and eight distinct lumped hydrological models were utilized. The evaluation of the method's performance was based on the CRPS (Continuous Ranked Probability Score), ABDU (Average Bin Distance to Uniformity) and NMIQR (Normalized Mean Interquartile Range) metrics.

The implementation of the three different configurations of Quantile

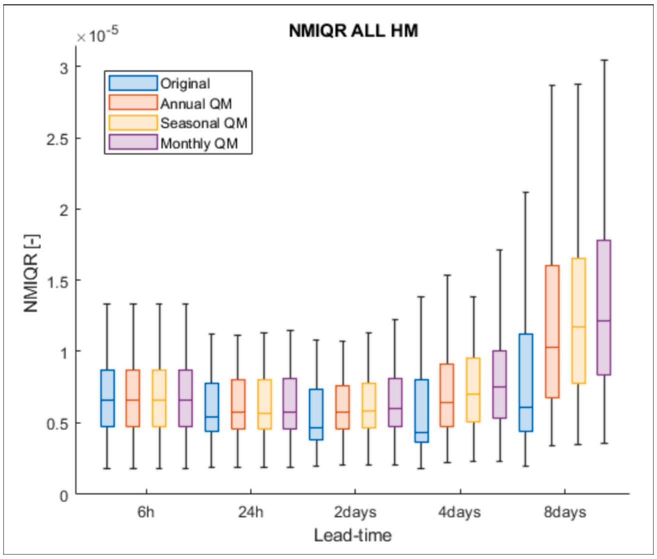

Fig. 11. Distributions of the normalized mean interquartile range (NMIQR) for streamflow forecasts from the four different precipitation inputs and 5 different lead-times (6 h, 24 h, 2 days, 4 days and 8 days). Each boxplot (using the 5th, 25th, 50th, 75th, and 95th percentiles) summarizes the ensemble of NMIQR scores over all 20 basins and all 8 hydrological models.

Mapping as a post-processing technique effectively reduced the CRPS values in precipitation forecasts in half of watersheds, which then reduce the CRPS in hydrological forecasts for the same half of the watersheds. However, a third of the catchments showed no effect and some specific hydrological models even displayed higher CRPS values at longer lead-times. In two cases, no improvement was observed, and the performance deteriorated. Despite these reductions in CRPS, it is important to note that the post-processing approach does not make the precipitation ensembles fully reliable, as indicated by poor improvements in the Talagrand diagrams and ABDU scores. Additionally, forecasts with QM precipitation tend to be less sharp than the raw forecasts, as shown by their higher NMIQR values. The CRPS, ABDU, and NMIQR metrics together provided a more comprehensive assessment of the overall quality of ensemble forecasts, highlighting areas of improvement and remaining challenges.

Furthermore, even if the study revealed the hydrological model GR5dt superiority over the others but only with a 10 to 30 % lower median CRPS compared to the rest of hydrological models. The investigation employed eight different hydrological models: CEQUEAU, GR5dt, HBV, HYMOD, IHACRES, MOHYSE, SIMHYD, and TOPMODEL. All these models exhibited acceptable performance relative to the catchments utilized in this study.

The study also identified several limitations and areas for future research. Firstly, the inclusion of more catchments in different locations with diverse hydroclimatic characteristics in Canada is necessary to obtain a comprehensive analysis. Moreover, employing model ensembles is crucial due to the varying performance of hydrological models across catchments in simulating sub-daily streamflow.

Additionally, the evaluation of hydrological models should involve the use of multiple performance metrics, considering factors such as flow variability, heteroscedastic errors, benchmark models, and specific model applications. Furthermore, the potential benefits of both weather and streamflow post-processing methods should be considered, therefore, future research should assess the combined impact of these techniques on hydrological forecasting accuracy.

In summary, while the Quantile Mapping method showed improvements in hydrological forecasts in this study, addressing the identified limitations and exploring more advanced post-processing techniques and models would be crucial for enhancing the accuracy and reliability of hydrological forecasts in diverse catchment areas.

Credit authorship contribution statement

Freya Saima Aguilar Andrade: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Richard Arsenault:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Annie Poulin:** Writing – review & editing, Validation, Supervision, Funding acquisition. **Magali Troin:** Writing – review & editing, Validation. **William Armstrong:** Software, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank the data support team at Water Survey Canada and Environment and Climate Change Canada for providing the sub-daily streamflow observation data. In this study, the ERA5 reanalysis dataset produced by Hersbach et al., (2018) was used. It has been downloaded from the Copernicus Climate Change Service (C3S) Climate Data Store: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview. The authors are also grateful to the European Centre for Medium-range Weather Forecasts (ECMWF) for providing access to archived IFS forecast data from their MARS computing and archiving facilities. The basemap in F ig. 1 was created using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. Copyright © Esri. All rights reserved. For more information about Esri® software, please visit www.esri.com.

Last but not least, the authors would like to thank the anonymous reviewers whose comments and suggestions helped improve the manuscript and shape it into its current form.

Funding

This study was partially funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) under the Alliance grant number ALLRP 560780–20.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2024.131861.

References

Andraos, C., & Najem, W. (2020). Multi-model approach for reducing uncertainties in rainfall-runoff models. In Advances in Hydroinformatics: SimHydro 2019-Models for Extreme Situations and Crisis Management (pp. 545-557). Springer Singapore.

Arsenault, R., Poulin, A., Côté, P., Brissette, F., 2014. A comparison of stochastic optimization algorithms in hydrological model calibration. J. Hydrol. Eng. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000938.

Arsenault, R., Gatien, P., Renaud, B., Brissette, F., Martel, J.L., 2015. A comparative analysis of 9 multi-model averaging approaches in hydrological continuous streamflow simulation. Journal of Hydrology 529, 754–767.

Arsenault, R., Latraverse, M., Duchesne, T., 2016. An efficient method to correct under-dispersion in ensemble streamflow prediction of inflow volumes for reservoir optimization. Water Resources Management 30, 4363–4380.
Arsenault, R., Brissette, F., Martel, J.L., 2018. The hazards of split-sample validation in hydrological model calibration. Journal of Hydrology 566, 346–362.
Baker, S.A., Wood, A.W., Rajagopalan, B., 2019. Developing subseasonal to seasonal climate forecast products for hydrology and water management. JAWRA Journal of the American Water Resources Association 55 (4), 1024–1037.
Bennett, A., Stein, A., Cheng, Y., Nijssen, B., McGuire, M., 2022. A Process-Conditioned and Spatially Consistent Method for Reducing Systematic Biases in Modeled Streamflow. J. Hydrometeor. 23, 769–783. https://doi.org/10.1175/JHM-D-21-0174.1.
Berg, P., Haerter, J., 2013. Unexpected increase in precipitation intensity with temperature—A result of mixing of precipitation types? Atmospheric Research 119, 56–61.
Bergeron, J., Leconte, R., Trudel, M., Farhoodi, S., 2021. On the Choice of Metric to Calibrate Time-Invariant Ensemble Kalman Filter Hyper-Parameters for Discharge Data Assimilation and Its Impact on Discharge Forecast Modelling. Hydrology 8 (1), 36.
Bergström, S., Forsman, A., 1973. Development of a conceptual deterministic rainfall-runoff mode. Nord. Hydrol 4, 240–253.
Beven, K.J., Kirkby, M.J., Schofield, N., Tagg, A.F., 1984. Testing a physically-based flood forecasting model (TOPMODEL) for three UK catchments. Journal of Hydrology 69 (1–4), 119–143.
Biemans, H., Hutjes, R.W.A., Kabat, P., Strengers, B.J., Gerten, D., Rost, S., 2009. Effects of precipitation uncertainty on discharge calculations for main river basincatchments. Journal of Hydrometeorology 10 (4), 1011–1025.
Boucher, M.A., Perreault, L., Anctil, F., Favre, A.C., 2015. Exploratory analysis of statistical post-processing methods for hydrological ensemble forecasts. Hydrological Processes 29 (6), 1141–1155.
Bourgin, F., Ramos, M.H., Thirel, G., Andréassian, V., 2014. Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting. Journal of Hydrology 519, 2775–2784.
Chen, J., Brissette, F.P., Zhang, X.J., Chen, H., Guo, S., Zhao, Y., 2019. Bias correcting climate model multi-member ensembles to assess climate change impacts on hydrology. Climatic Change 153, 361–377.
Chiew, F.H.S., Peel, M.C., Western, A.W., 2002. Application and testing of the simple rainfall-runoff model SIMHYD. Mathematical Models of Small Watershedcatchment Hydrology and Applications 335–367.
Copernicus Products (2018). ERA5 hourly data on single levels from 1979 to present, ECMWF: https://cds.climate.copernicus.eu/.
Dale, M., 2021. Managing the effects of extreme sub-daily rainfall and flash floods—a practitioner's perspective. Philosophical Transactions of the Royal Society A 379 (2195), 20190550.
Dion, P., Martel, J.L., Arsenault, R., 2021. Hydrological ensemble forecasting using a multi-model framework. Journal of Hydrology 600, 126537.
Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. Ocean Dynamics 53, 343–367. https://doi.org/10.1007/s10236-003-0036-9.
Fang, G.H., Yang, J., Chen, Y.N., Zammit, C., 2015. Comparing bias correction methods in downscaling meteorological variables for a hydrologic impact study in an arid area in China. Hydrology and Earth System Sciences 19 (6), 2547–2559.
Ferreira, P.M.D.L., Paz, A.R.D., Bravo, J.M., 2020. Objective Functions Used as Performance Metrics for Hydrological Models: State-of-the-Art and Critical Analysis. RBRH, 25.
Ficchi, A., 2017. An adaptive hydrological model for multiple time-steps: Diagnostics and improvements based on fluxes consistency. UPMC -, Irstea Antony, Paris, France. PhD thesis.
Ficchi, A., Perrin, C., Andréassian, V., 2019. Hydrological modelling at multiple sub-daily time steps: model improvement via flux-matching. Journal of Hydrology 575, 1308–1327. https://doi.org/10.1016/j.jhydrol.2019.05.084.
Fortin, V., Turcotte, R., 2007. Le modèle hydrologique MOHYSE. Université du Québec à Montréal, Montreal, Canada, Département des sciences de la Terre et de l'atmosphère.
Ghimire, U., Srinivasan, G., Agarwal, A., 2019. Assessment of rainfall bias correction techniques for improved hydrological simulation. International Journal of Climatology 39 (4), 2386–2399.
Girard, G., Morin, G., Charbonneau, R., 1972. Modèle précipitations-débits à discrétisation spatiale. Cahiers ORSTOM, Série Hydrologie 9 (4), 35–52.
Graybeal, D.Y., Leathers, D.J., 2006. Snowmelt-related flood risk in Appalachia: First estimates from a historical snow climatology. Journal of Applied Meteorology and Climatology 45 (1), 178–193.
Guerreiro, S.B., Fowler, H.J., Barbero, R., Westra, S., Lenderink, G., Blenkinsop, S., Lewis, E., Li, X.-F., 2018. Detection of continental-scale intensifcation of hourly rainfall extremes. Nature Climate Change 8 (9), 803–807.
Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377 (1–2), 80–91.
Hamill, T.M., 2001. Interpretation of rank histograms for verifying ensemble forecasts. Monthly Weather Review 129 (3), 550–560.
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Thépaut, J.N., 2020. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society 146 (730), 1999–2049.
Hopson, T.M., Wood, A.W., Weerts, A.H., 2018. Motivation and overview of hydrological ensemble post-processing. In: Handbook of Hydrometeorological Ensemble Forecasting. Springer Verlag. https://doi.org/10.1002/hyp.9409.
Jakeman, A.J., Littlewood, I.G., Whitehead, P.G., 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. Journal of Hydrology 117 (1–4), 275–300.
Jie, M.X., Chen, H., Xu, C.Y., Zeng, Q., Tao, X.E., 2016. A comparative study of different objective functions to improve the flood forecasting accuracy. Hydrology Research 47 (4), 718–735.
Katiraie-Boroujerdy, P.S., Rahnamay Naeini, M., Akbari Asanjan, A., Chavoshian, A., Hsu, K.L., Sorooshian, S., 2020. Bias correction of satellite-based precipitation estimations using quantile mapping approach in different climate regions of Iran. Remote Sensing 12 (13), 2102.
Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences 5, 89–97.
Leutbecher, M., Palmer, T.N., 2008. Ensemble forecasting. Journal of Computational Physics 227 (7), 3515–3539.
Li, W., Duan, Q., Miao, C., Ye, A., Gong, W., Di, Z., 2017. A review on statistical postprocessing methods for hydrometeorological ensemble forecasting. Wiley Interdisciplinary Reviews: Water 4 (6). https://doi.org/10.1002/wat2.1246.
Li, Y., Jiang, Y., Lei, X., Tian, F., Duan, H., Lu, H., 2018. Comparison of precipitation and streamflow correcting for ensemble streamflow forecasts. Water 10 (2), 177.
Li, X., Wu, H., Nanding, N., Chen, S., Hu, Y., Li, L., 2023. Statistical Bias Correction of Precipitation Forecasts Based on Quantile Mapping on the Sub-Seasonal to Seasonal Scale. Remote Sensing 15 (7), 1743.
Liu, Y., Weerts, A.H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A.I.J.M., van Velzen, N., He, M., Lee, H., Noh, S.J., Rakovec, O., Restrepo, P., 2012. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. Hydrology and Earth System Sciences 16 (10), 3863–3887. https://doi.org/10.5194/hess-16-3863-2012.
Liu, X., Zhang, L., She, D., Chen, J., Xia, J., Chen, X., Zhao, T., 2022. Postprocessing of hydrometeorological ensemble forecasts based on multisource precipitation in Ganjiang River basin. China. Journal of Hydrology 605, 127323.
Lucatero, D., Madsen, H., Refsgaard, J.C., Kidmose, J., Jensen, K.H., 2018. Seasonal streamflow forecasts in the Ahlergaarde catchment, Denmark: the effect of preprocessing and post-processing on skill and statistical consistency. Hydrology and Earth System Sciences 22 (7), 3601–3617.
Madadgar, S., Moradkhani, H., Garen, D., 2014. Towards improved post-processing of hydrologic forecast ensembles. Hydrological Processes 28 (1), 104–122.
Mai, J., 2023. Ten strategies towards successful calibration of environmental models. Journal of Hydrology 620, 129414.
Maraun, D., 2016. Bias correcting climate change simulations-a critical review. Current Climate Change Reports 2 (4), 211–220.
Matheson, J.E., Winkler, R.L., 1976. Scoring rules for continuous probability distributions. Management Science 22 (10), 1087–1096.
Mathevet, T., 2005. Which rainfall-runoff model at the hourly time-step? Empirical development and intercomparison of rainfall runoff model on a large sample of watershed catchments. French.] Ph. d. thesis, Cemagref, École nationale du génie rural, des eaux et des forêts Univ.
Maurer, E.P., Pierce, D.W., 2014. Bias correction can modify climate model simulated precipitation changes without adverse effect on the ensemble mean. Hydrology and Earth System Sciences 18 (3), 915–925.
McInerney, D., Thyer, M., Kavetski, D., Laugesen, R., Woldemeskel, F., Tuteja, N., Kuczera, G., 2022. Seamless streamflow forecasting at daily to monthly scales: MuTHRE lets you have your cake and eat it too. Hydrol. Earth Syst. Sci. 26, 5669–5683. https://doi.org/10.5194/hess-26-5669-2022.
McMillan, H., Jackson, B., Clark, M., Kavetski, D., Woods, R., 2011. Rainfall uncertainty in hydrological modelling: An evaluation of multiplicative error models. Journal of Hydrology 400 (1–2), 83–94.
McMillan, H., Krueger, T., Freer, J., 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrological Processes 26 (26), 4078–4111.
Miao, C., Su, L., Sun, Q., Duan, Q., 2016. A nonstationary bias-correction technique to remove bias in GCM simulations. Journal of Geophysical Research: Atmospheres 121 (10), 5718–5735.
Moriasi, D.N., Gitau, M.W., Pai, N., Daggupati, P., 2015. Hydrologic and water quality models: Performance measures and evaluation criteria. Transactions of the ASABE 58 (6), 1763–1785.
Muhammad, A., Evenson, G.R., Stadnyk, T.A., Boluwade, A., Jha, S.K., Coulibaly, P., 2019. Impact of model structure on the accuracy of hydrological modeling of a Canadian Prairie watershed. Journal of Hydrology: Regional Studies 21, 40–56. https://doi.org/10.1016/j.ejrh.2018.11.005.
Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I—A discussion of principles. Journal of Hydrology 10 (3), 282–290.
Niranjan Kumar, K., Thota, M.S., Ashrit, R., Mitra, A.K., Rajeevan, M.N., 2022. Quantile mapping bias correction methods to IMDAA reanalysis for calibrating NCMRWF unified model operational forecasts. Hydrological Sciences Journal 67 (6), 870–885.
Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., Loumagne, C., 2005a. Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. Journal of Hydrology 303 (1–4), 290–306.
Oudin, L., Michel, C., Anctil, F., 2005b. Which potential evapotranspiration input for a lumped rainfall-runoff model?: Part 1—Can rainfall-runoff models effectively handle detailed potential evapotranspiration inputs? Journal of Hydrology 303 (1–4), 275–289.
Pagano, T.C., Wood, A.W., Ramos, M., Cloke, H.L., Pappenberger, F., Clark, M.P., et al., 2014. Challenges of operational river forecasting. Journal of Hydrometeorology 15, 1692–1707. https://doi.org/10.1175/JHM-D-13-0188.1.

P Perrin, C., Michel, C., Andréassian, V., 2003. Improvement of a parsimonious model for streamflow simulation. J. Hydrol. 279, 275–289. 10.1016/S0022-1694(03)00225-7.
Piazzi, G., et al., 2021. Sequential data assimilation for streamflow forecasting: assessing the sensitivity to uncertainties and updated variables of a conceptual hydrological model at basin scale. Water Resources Research 57 (4).
Pomeroy, J.W., Stewart, R.E., Whitfield, P.H., 2016. The 2013 flood event in the South Saskatchewan and Elk River basins: Causes, assessment and damages. Canadian Water Resources Journal/revue Canadienne Des Ressources Hydriques 41 (1–2), 105–117.
Pushpalatha, R., Perrin, C., Le Moine, N., Andréassian, V., 2012. A review of efficiency criteria suitable for evaluating low-flow simulations. Journal of Hydrology 420, 171–182.
Rayner, S., Lach, D., Ingram, H., 2005. Weather forecasts are for wimps: why water resource managers do not use climate forecasts. Climatic Change 69 (2), 197–227.
Robertson, D.E., Shrestha, D.L., Wang, Q.J., 2013. Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. Hydrology and Earth System Sciences 17 (9), 3587–3603.
Seiller, G., Anctil, F., 2016. How do potential evapotranspiration formulas influence hydrological projections? Hydrological Sciences Journal 61 (12), 2249–2266.
Shastri, H., Ghosh, S., Karmakar, S., 2017. Improving Global Forecast System of extreme precipitation events with regional statistical model: Application of quantile-based probabilistic forecasts. Journal of Geophysical Research: Atmospheres 122 (3), 1617–1634.
Shaw, D.A., Pietroniro, A.L., Martz, L.W., 2013. Topographic analysis for the prairie pothole region of Western Canada. Hydrological Processes 27 (22), 3105–3114.
Shen, H., Tolson, B.A., Mai, J., 2022. Time to update the split-sample approach in hydrological model calibration. Water Resources Research 58 (3).
Tani, S., Gobiet, A., 2019. Quantile mapping for improving precipitation extremes from regional climate models. Journal of Agrometeorology.
Tao, Y., Duan, Q., Ye, A., Gong, W., Di, Z., Xiao, M., Hsu, K., 2014. An evaluation of post-processed TIGGE multimodel ensemble precipitation forecast in the Huai river basin. Journal of Hydrology 519, 2890–2905.
Tarek, M., Brissette, F. P. y Arsenault, R. (2020) Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. Hydrology and Earth System Sciences, 24, pp. 2527–2544.
Thiboult, A., Anctil, F., 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.
Trinh-Tuan, L., Matsumoto, J., Tangang, F.T., Juneng, L., Cruz, F., Narisma, G., Ngo-Duc, T., 2019. Application of quantile mapping bias correction for mid-future precipitation projections over Vietnam. Sola 15, 1–6.
Troin, M., Poulin, A., Baraer, M., Brissette, F., 2016. Comparing snow models under current and future climates: Uncertainties and implications for hydrological impact studies. Journal of Hydrology 540, 588–602.
Troin, M., Arsenault, R., Wood, A.W., Brissette, F., Martel, J.L., 2021. Generating ensemble streamflow forecasts: A review of methods and approaches over the past 40 years. Water Resources Research 57. https://doi.org/10.1029/2020WR028392.
Valéry, A., 2010. Modélisation précipitations débit sous influence nivale: Elaboration d'un module neige et évaluation sur 380 bassins versants. Thèse de Doctorat, Cemagref (Antony), AgroParisTech, Paris.
Velazquez, J.A., Anctil, F., Ramos, M.H., Perrin, C., 2011. Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. Advances in Geosciences 29, 33–42.
Verkade, J.S., Brown, J.D., Reggiani, P., Weerts, A.H., 2013. Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. Journal of Hydrology 501, 73–91.
Wagener, T., Boyle, D.P., Lees, M.J., Wheater, H.S., Gupta, H.V., Sorooshian, S., 2001. A framework for development and application of hydrological models. Hydrology and Earth System Sciences 5 (1), 13–26.
Wilcke, R.A.I., Mendlik, T., Gobiet, A., 2013. Multi-variable error correction of regional climate models. Climatic Change 120, 871–887.
Wu, L., Seo, D.J., Demargne, J., Brown, J.D., Cong, S., Schaake, J., 2011. Generation of ensemble precipitation forecast from single-valued quantitative precipitation forecast for hydrologic ensemble prediction. Journal of Hydrology 399 (3–4), 281–298.
Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. Journal of Hydrology 181 (1–4), 23–48.
Zhang, Z. (2015) Data Assimilation. Mathematical and Physical Fundamentals of Climate Change :291-311. 10.1016/B978-0-12-800066-3.00009-7.