



# Verification of ECMWF System 4 for seasonal hydrological forecasting in a northern climate

Rachel Bazile<sup>1,\*</sup>, Marie-Amélie Boucher<sup>1</sup>, Luc Perreault<sup>2</sup>, and Robert Leconte<sup>1</sup>

<sup>1</sup>Département de génie civil, Université de Sherbrooke, 2500 Boul. de l'Université, Sherbrooke, Québec, J1R 2R2, Canada

<sup>2</sup>Institut de Recherche d'Hydro-Québec (IREQ), 1800 boul. Lionel-Boulet, Varennes, Québec, J3X 1S1, Canada

\* *Invited contribution by Rachel Bazile, recipient of the EGU Hydrological Sciences Outstanding Student Poster and PICO Award 2017.*

Correspondence to: Rachel Bazile (rachel.bazile@gadz.org)

Received: 30 June 2017 – Discussion started: 12 July 2017

Revised: 26 September 2017 – Accepted: 30 September 2017 – Published: 22 November 2017

**Abstract.** Hydropower production requires optimal dam and reservoir management to prevent flooding damage and avoid operation losses. In a northern climate, where spring freshet constitutes the main inflow volume, seasonal forecasts can help to establish a yearly strategy. Long-term hydrological forecasts often rely on past observations of streamflow or meteorological data. Another alternative is to use ensemble meteorological forecasts produced by climate models. In this paper, those produced by the ECMWF (European Centre for Medium-Range Forecast) System 4 are examined and bias is characterized. Bias correction, through the linear scaling method, improves the performance of the raw ensemble meteorological forecasts in terms of continuous ranked probability score (CRPS). Then, three seasonal ensemble hydrological forecasting systems are compared: (1) the climatology of simulated streamflow, (2) the ensemble hydrological forecasts based on climatology (ESP) and (3) the hydrological forecasts based on bias-corrected ensemble meteorological forecasts from System 4 (corr-DSP). Simulated streamflow computed using observed meteorological data is used as benchmark. Accounting for initial conditions is valuable even for long-term forecasts. ESP and corr-DSP both outperform the climatology of simulated streamflow for lead times from 1 to 5 months depending on the season and watershed. Integrating information about future meteorological conditions also improves monthly volume forecasts. For the 1-month lead time, a gain exists for almost all watersheds during winter, summer and fall. However, volume forecasts

performance for spring varies from one watershed to another. For most of them, the performance is close to the performance of ESP. For longer lead times, the CRPS skill score is mostly in favour of ESP, even if for many watersheds, ESP and corr-DSP have comparable skill. Corr-DSP appears quite reliable but, in some cases, under-dispersion or bias is observed. A more complex bias-correction method should be further investigated to remedy this weakness and take more advantage of the ensemble forecasts produced by the climate model. Overall, in this study, bias-corrected ensemble meteorological forecasts appear to be an interesting source of information for hydrological forecasting for lead times up to 1 month. They could also complement ESP for longer lead times.

## 1 Introduction

Hydropower production planning typically requires inflow forecasts for reservoirs at different lead times. Whereas short-term forecasts are used for day-to-day planning, sub-seasonal (1 to 3 months) to seasonal (up to 6 month) forecasts are used to establish a yearly strategy. Improving the skill of hydrological forecasts at the sub-seasonal up to seasonal scale is thus essential. According to DelSole (2004), “a system is said to be unpredictable if the forecast distribution, which gives the most complete description of the future state based on all available knowledge, is identical to the climato-

logical distribution, which describes the state in the absence of time lag information". Hence, the advent of seasonal meteorological forecasts that are more informative than climatology could support water managers in their decision-making process.

Probabilistic forecasts are necessary to quantify uncertainty about future hydrological conditions. It is especially true for long-term forecasts, as uncertainty grows with longer lead times. Operationally, several methods exist to produce sub-seasonal to seasonal hydrological forecasts. They can be broadly divided into two main categories: statistical forecasting and ensemble-based forecasting (Yuan et al., 2015). However, hybrid methods also exist.

Statistical methods can take advantage of relationships between past and future streamflow persistence (e.g. Svensson, 2016) or between streamflows and teleconnections indices. Examples include multiple regression type of models treated in a Bayesian perspective (e.g. Wang et al., 2009) or a frequentist framework (e.g. Moradkhani and Meier, 2010; Sveinsson et al., 2008).

Ensemble-based forecasting is a widespread uncertainty assessment technique (Cloke and Pappenberger, 2009). Ensemble forecasts comprise different potential future scenarios also called "members". One possible method to obtain hydrological ensembles is to provide ensemble meteorological forecasts as inputs to one or several hydrological models. Each meteorological scenario leads to one hydrological scenario (member). From these members, different techniques exist to derive probabilistic forecasts (Bröcker and Smith, 2008).

For long forecasting horizons, the simplest type of ensemble forecasts is the climatology of streamflow, either observed or simulated, hereafter called "historical streamflow prediction" (HSP). This naive forecasting method is by definition a reliable forecasting system, but of course its resolution can be improved. Even if this kind of forecasting system does not show any predictability, it accounts for different plausible hydrological scenarios based on the past. A simple alternative method, proposed by Day (1985), is called "extended streamflow prediction" (ESP). To produce ESP, past meteorological observations are considered as equiprobable potential future meteorological scenarios. Contrary to HSP, ESP requires the use of a calibrated hydrological model to produce hydrological forecasts. If the historical record is long enough, climatology provides a reliable estimation of the distribution of future meteorological conditions, including some extreme scenarios. The main advantage of ESP relative to HSP is that it allows accounting for the current hydrological initial conditions. Several studies have shown that state variables such as soil moisture or snow water equivalent can provide relevant information to extend predictability for lead times from one to several months ahead (e.g. Wood and Lettenmaier, 2008; Shukla et al., 2013; Yang et al., 2014; Yuan et al., 2016). The influence of initial conditions depends on the period of the year and on the location of watersheds

(e.g. Yossef et al., 2013). ESP is intuitively appealing, since it is coherent with a natural tendency of humans to judge actual situations according to their memory of past experiences. Moreover, ESP allows practitioners to condition streamflow scenarios only on selected past meteorological scenarios if they wish, for instance by using only the most extremes historical scenarios. Because of their simplicity and efficiency, both ESP and HSP are popular among operational agencies for forecasts from several days to weeks to months in the future (e.g. García-Morales and Dubus, 2007) and still arouse interest as a forecasting system (e.g. Singh, 2016).

However, given the current context of climate change, some past meteorological and hydrological data might not be representative of plausible future conditions. In northern contexts, it is expected that climate change will gradually modify the repartition of rain and snowfall during the year, for instance. For the province of Québec in Canada, climatic projections anticipate a rise in temperature and precipitation (Ouranos, 2015). It is expected that these changes will modify hydrological conditions both at the annual and intra-annual scales. Indeed, higher winter streamflows, earlier spring freshet and longer periods of low streamflow during the summer are expected (Guay et al., 2015). In central Sweden, climate change will also affect the seasonality of streamflow, mostly by decreasing the mean snow water equivalent and the mean annual runoff (Xu, 2000).

During the past decade, sub-seasonal to seasonal ensemble meteorological forecasts produced by dynamic climate models have undergone constant improvements, and it is worth assessing their usefulness for long lead-time inflow forecasting. A dynamic climate model is an atmospheric model, sometimes coupled with an ocean model. Considering the interactions between the atmosphere and ocean allows for modelling long-term phenomenon such as El Niño and La Niña phases of the El Niño–Southern Oscillation cycle. For instance, according to Kim et al. (2012), for winter in the Northern Hemisphere, both the European Centre for Medium-Range Weather Forecasts (ECMWF) System 4 and the National Centers for Environmental Prediction (NCEP) Climate Forecast System Version 2 (CFSv2) accurately reproduce El Niño/La Niña phases. Temperature variations are more difficult to capture. Regarding the ECMWF System 4, Weisheimer and Palmer (2014) assessed the performance of 2 m temperature and precipitation forecasts throughout the world. The reliability of forecasts vary from "perfect" to "dangerous", depending on the month of the year, the variables and the location.

Meteorological forecasts from dynamic climate models can be used to produce hydrological forecasts, hereafter called "dynamical streamflow prediction" (DSP). However, according to previous studies, their potential for hydrological purposes is highly variable, depending on the location and the context. Luo and Wood (2008) compared forecasts from the NCEP Climate Forecast System (CFS), multimodel forecasts from a combination of CFS and seven models from

the DEMETER (Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction; Palmer et al., 2004) database and ESP for hydrological forecasting on a watershed in Ohio during summer. They found that the multimodel approach is more efficient than a single climate model. Both outperform the ESP approach in terms of ranked probability score (CRPS). Multimodel approaches improve seasonal forecasts significantly for the 1- and 2-month lead times, whereas single model improvements are limited. Across the United States, the ESP approach has been compared to hydrological ensemble forecasts based on NCEP CFSv1 and CFSv2 by Yuan et al. (2013). Their results indicate that CFSv2 improves hydrological forecasting performances for the 1-month lead time, whereas CFSv1-based forecasts are not very efficient. Similarly, He et al. (2016) compared the performance of climatology (ESP) and CFSv2 for a single watershed in the Sierra Nevada. Their results indicate only little improvement when using CFSv2. Some agencies already integrate information from long-term ensemble forecasts into their operational hydrological forecasting systems (e.g. Demargne et al., 2014). One major problem of DSP is that ensemble meteorological forecasts produced by dynamic models suffer from bias. However, in the context of climate change, using dynamical meteorological forecasts seems intuitively valuable, as they are expected to better represent the current climate, compared to methods based on past climatology (such as ESP). Crochemore et al. (2016) compared different strategies for bias correction of daily precipitation forecasts and evaluated their efficiency for hydrological forecasting over 16 watersheds in France. They show that correcting precipitation forecasts does indeed translate into an improvement of hydrological forecasts. However, they also show that simple bias-correction methods, such as linear scaling, are as efficient as more sophisticated methods.

The goal of this study is to evaluate the potential of DSP in terms of predictability improvement for long-term streamflow forecasting, compared to HSP and ESP. More specifically, long-term meteorological and hydrological forecasts are assessed for 10 northern watersheds in the province of Québec in Canada. Those watersheds are all exploited for hydropower production. Therefore, skillful long lead-time forecasts are crucial for optimal water management, especially for anticipating and exploiting the large inflow to reservoirs during spring melt.

The paper is organised as follow. After describing the context in Sect. 2, details regarding the case studies and available data will be given in Sect. 3. The forecast verification methodology is presented in Sect. 4. Results are presented and discussed in Sect. 5, and conclusions are drawn in Sect. 6, which also identifies potential research avenues for future studies.

## 2 Hydro-meteorological context of the study

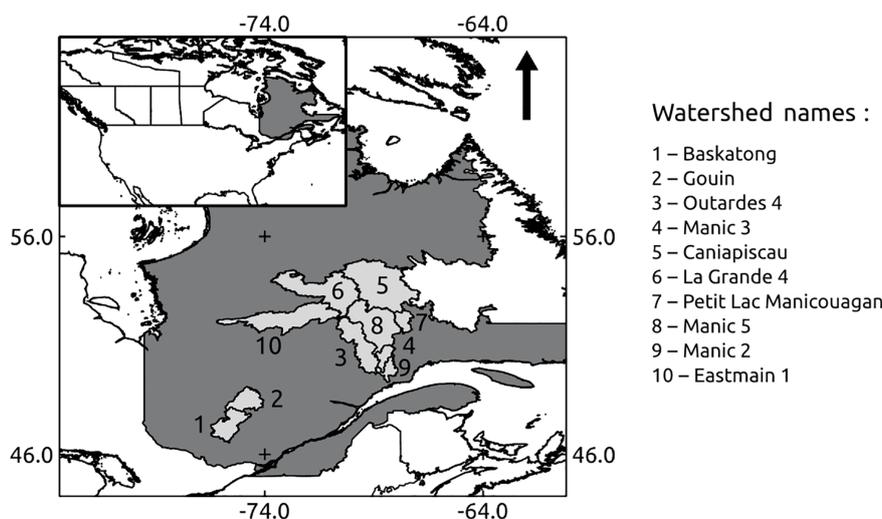
### 2.1 Watersheds

The 10 watersheds used as a test bed in this study are all located in the province of Québec and exploited by Hydro-Québec to generate hydropower. Hydro-Québec is a government-owned corporation that produces and distributes electricity in the province of Québec. The installed hydroelectricity capacity of Hydro-Québec is more than 36 000 MW. Together, the 10 watersheds under study represent more than 8750 MW (Hydro-Québec, 2015), as the outlet of each watershed is a hydropower reservoir. Figure 1 illustrates the geographical location of the 10 watersheds. Two of them are located in the southern portion of the province, whereas the others are located in the central portion. In addition, some watersheds are part of larger hydropower production complexes. For instance, watersheds 4, 7, 8 and 9 comprise the Manicouagan complex and watersheds 5, 6 and 10 are part of the La Grande complex.

The streamflow regime of the 10 watersheds is dominated by a northern climate, which induces snow accumulation and low streamflow during winter (December to February), followed by high streamflow during spring. The exact timing of the spring freshet for a particular watershed is a function of not only its latitude and physiographic characteristics, such as slope and orientation, but also of the meteorological conditions that prevail during a particular year. Watersheds located in the southern portion of the province generally produce their highest streamflows in March or April, whereas those located in the central and northern parts reach their maximum streamflows in May or June. The total volume of runoff associated with the spring freshet obviously depends on the accumulated snow pack during the winter season. Most watersheds also exhibit high streamflow during fall on most years, when evapotranspiration is low and soil is saturated. In fact, whereas temperatures generally show large variability during the winter, the spatiotemporal variability of precipitation is higher during summer and fall. Table 1 presents the hydro-meteorological characteristics of the 10 watersheds under study.

### 2.2 Current operational streamflow forecasting system

The current operational streamflow forecasting system at Hydro-Québec relies on ESP (Day, 1985) and can be divided into three distinct stages. In the first stage, an analogue approach (e.g. Marty et al., 2012) based upon deterministic meteorological forecasts from Environment and Climate Change Canada is used to produce short-term meteorological forecasts. The definition of “short term” is not fixed. The lead time depends on watersheds and meteorological events. On average, 5 to 7 days ahead forecasts are produced using this method. The second stage aims to produce seasonal forecasts. Observed precipitation and temperature for previ-



**Figure 1.** Geographical location of the watersheds used in this study.

**Table 1.** Hydro-meteorological characteristics of the watersheds illustrated in Fig. 1

ID (see Fig. 1)	Area (km <sup>2</sup> )	Annual min. temp.* (°C)	Annual mean temp.* (°C)	Annual max. temp.* (°C)	Mean annual precip. (mm)	Annual min. streamflow (m <sup>3</sup> s <sup>-1</sup> )	Annual mean streamflow (m <sup>3</sup> s <sup>-1</sup> )	Annual max. streamflow (m <sup>3</sup> s <sup>-1</sup> )	Mean date of the max. spring streamflow (day/month)	Proportion of the spring freshet volume compared to the annual volume (%)
1	13 057	-30	2.2	23.9	1027	17	231	1306	28/04	45
2	9 426	-31.2	1.3	23.8	980	11	180	1282	11/05	40
3	17 119	-33.6	-1.2	21.7	918	80	359	1759	20/05	44
4	4 245	-31.1	0.1	21.6	980	19	104	536	18/05	45
5	37 328	-34.7	-3.6	20.5	807	151	747	3073	30/05	38
6	28 443	-36.6	-3.4	20.9	816	140	564	1392	09/06	30
7	4 565	-33.3	-2.3	20.9	909	12	106	638	25/05	45
8	24 608	-33.6	-2.0	20.8	910	41	526	3276	21/05	48
9	4 100	-28	1.5	21.5	1015	5	77	537	15/05	51
10	26 944	-35.8	-2.5	21.2	848	109	572	1981	23/05	37

\* Based on the mean daily temperatures.

ous years are considered as plausible future scenarios. Hence, archived observed meteorological conditions for all previous years (since 1950) form an ensemble. These analogue-based meteorological ensembles are used to extend the short-term forecasts obtained in the first stage. These scenarios are then fed to a lumped conceptual hydrological model (described below) to obtain hydrological ensemble forecasts. Lastly, the third and last stage begins when the influence of initial conditions becomes negligible. Observed streamflow for the same Julian day of each available year in the database are then considered as equiprobable long-term forecasts (HSPs; see Introduction). The appropriate moment to shift from ESP to HSP is fixed by the forecaster and varies between watersheds. Note that Hydro-Québec is currently improving its forecasting system by integrating ensemble weather forecasts with statistical post-processing for short-term forecasting, and by developing a weather generator for medium-term forecast-

ing. This new system is expected to become operational in 2018.

The available archive of past meteorological observations covers the 1950–2015 period. Data include daily minimum and maximum temperature as well as daily rainfall and snowfall. Those variables are only available at the watershed scale, meaning that observations from individual weather stations were spatially aggregated before being archived. Those weather stations are part of a province-wide cooperative network called RMCQ (in French *Réseau météorologique coopératif du Québec*; Lepage and Bourgeois, 2011). The aim of this cooperative network is to pool together data from private and public collaborators. Unfortunately, the number of stations and the interpolation method have evolved over time. At the time of writing, it was not possible to obtain detailed information regarding those successive changes. However, meteorological data are generally of good quality and

there are no missing dates. Daily data are collected from 06:00 to 06:00 UTC for precipitation and temperature and from 05:00 to 05:00 UTC for streamflow data.

For the purpose of this study, climatology-based ensemble forecasts were built. All available years were used except one, in rotation. For instance, the climatology-based forecasts for year 1980 include all years but 1980. This setup is of course different than the operational framework, where the ensemble size grows year after year and the future is unknown. However, the methodology used here allows one to maintain a constant ensemble size (64 members). In addition, this produces ensembles that are free of any possible trend in the time series of climate data. Indeed, all information about past meteorological conditions are used as inputs to the hydrological model.

In all cases, meteorological series are used as inputs to HSAMI, a lumped conceptual hydrological model described below (in French, Fortin, 2000). This model is based on a series of three linear reservoirs which supply two hydrographs. Snow accumulation and melt are based on a degree-day approach. HSAMI uses daily minimum and maximum temperature, as well as rainfall and snowfall to compute the mean streamflow at the outlet of the watershed at a daily time step. The model has 23 parameters that must be calibrated against previous streamflow observations. The sets of parameters used in this study are provided by Hydro-Québec. Modelling performance varies greatly from one watershed to another. The Nash–Sutcliffe efficiencies (NSEs) based on daily streamflow data range from 0.30 to 0.86 for the 1981–2015 period. Despite low NSEs for some watersheds, it was judged appropriate to use these parameter sets rather than recalibrating the model. First, this variation is attributable mostly to the quality of hydrological data collected before 2000, on which the calibration of the model is based. Second, since the goal of this study is to assess the influence of meteorological forecasts on hydrological forecasts, simulated streamflows are used as pseudo “observations” in the verification process, and therefore a perfectly well-calibrated model is not required (see Sect. 4 for details).

The next section describes the ECMWF System 4 that is explored in this study as a potential replacement for the current operational forecasting system.

### 3 An alternative system for seasonal forecasting based on long-term dynamical climate modelling: the ECMWF System 4

Our hypothesis is that exploiting dynamical meteorological forecasts would improve the streamflow forecasting chain compared to ESP. The rationale behind this hypothesis is that dynamical meteorological forecasts should be driven by the current state of the atmosphere at their initialization. Eventually, considering the context of an evolving climate, this could also help hydropower producers to adapt reservoir and

dam management to new situations. The long-range ensemble meteorological forecasts used in this study are produced by the ECMWF (European Centre for Medium-Range Forecast) System 4.

System 4 (Molteni et al., 2011) is a global coupled ocean–atmosphere model that officially became operational in 2011. It is used to produce reforecasts and real-time forecasts, which are both archived. The atmospheric model component, namely the ECMWF IFS (Integrated Forecast System) model (version 36r4) includes a lake model and also involves ozone, volcanic aerosol and solar cycle action. Sea ice is depicted by initial sea-ice conditions for short lead times and by observed conditions for the 5 previous years. The initialization of the atmospheric model is performed using ERA-Interim for the reforecasts and the operational procedure of the ECMWF for real-time forecasts. The ocean model is initialised by the NEMOVAR ocean analysis. Ensemble forecasts are produced by perturbing initial conditions. In the current model setup, five members originate from perturbations of ocean wind surface initial conditions, whereas other members originate from sea surface temperature perturbations and stochastic physics. More details can be found on the ECMWF website (ECMWF, 2017).

Real-time forecasts are issued on the first day of each month for the next 215 days (approx. 7 months). Ensemble forecasts are computed at finer time steps than 1 day but are available only at daily time step from 00:00 to 00:00 UTC for this study. They are archived and available from 2012 to 2015. A set of reforecasts is also available, from 1981 to 2011. Reforecasts for the months of February, May, August and November as well as real-time forecasts comprise 51 members. Reforecasts for the other months comprise 15 members. Both archived past real-time forecasts and reforecasts are used in the present study. This allows the extension of the verification database length, but poses certain challenges in terms of performance assessment, since the number of members vary. In the following, the term “forecasts” will refer indifferently to real-time forecasts and reforecasts.

In the context of this study, the meteorological variables of interest are those that are inputs to HSAMI, namely daily minimum and maximum temperature as well as total daily precipitation. The original output grid of System 4 has a  $0.7^\circ$  horizontal resolution for the atmospheric model and around  $1^\circ$  for the ocean model at mid-latitudes. Since HSAMI is a lumped model, meteorological forecasts have to be a single point representative of the meteorological conditions over the watershed. The original resolution is too coarse for hydrological applications, as only very few grid points fall inside the watershed delineations. The original grid was thus down-scaled to a  $0.1^\circ$  grid through linear interpolation in order to obtain more points inside the watersheds boundaries. This allows ensuring that points close to the watersheds boundaries contribute to more accurate meteorological forecasts over the watershed. Then, grid points were averaged to aggregate the

information at the watershed scale. Total precipitation was separated into rainfall and snowfall according to air temperature.

#### 4 Forecast quality assessment

Both meteorological and hydrological forecasts are available at daily time steps. However, as mentioned in Sects. 2 and 3, a lag exists between daily forecasts and observations. A monthly aggregation of the different variables is performed for verification, in order to limit the impact of the lag between forecasts (meteorological and hydrological) and observations. Moreover, as the main goal of long-term hydro-meteorological forecasts is to provide information for seasonal to yearly dam management and optimisation of electricity production, decision makers are generally interested in inflow volumes for reservoirs. Consequently, monthly aggregated variables are considered. Monthly averages are computed for minimum and maximum temperatures. For precipitation and streamflow, monthly cumulative values are considered using calendar months. Many other types of information derived from streamflow forecasts are useful for dam management and reservoir operation. Anticipating runoff volume for spring freshet is crucial, as it allows for planning the lowering of the reservoirs to avoid risks of spillage and flooding. The inflow volume for the spring freshet is calculated between 1 March and 31 May for watersheds 1 and 2 and 1 April and 30 June for all other watersheds.

Both forecasts and reforecasts are pooled together to assess forecasts performance. Overall, 420 ensemble forecasts are available for verification purposes, as one ensemble forecast is issued on the first of each month between 1981 and 2015 (12 months over 35 years). Because both meteorological and streamflow observations are not available after 31 December 2015, lead times of 2 to 7 months have 419 to 413 forecast–observation pairs for the verification, respectively. The verification set should be as large as possible, in order to ensure statistical significance of the results. It should also be homogeneous. However, in reality, forecasts characteristics change depending on the period of the year and contradictory behaviours can balance each other out. Considering these two requirements, skill scores of monthly variables are calculated over the seasons. Four seasons are used, namely January–February–March (JFM), April–May–June (AMJ), July–August–September (JAS) and October–November–December (OND). For one season and one lead time, each set of verification comprises between 100 and 105 monthly ensemble forecast–observation pairs.

Different numerical scores and graphical tools are used to assess the quality of the aforementioned quantities. The joint use of several diagnostic devices is essential for different reasons. First, ensemble and probabilistic forecasts can be evaluated in terms of different attributes, and no single scoring rule can simultaneously assess them all. Second, exam-

ining different attributes can help to pinpoint strengths and weaknesses of competing forecasting systems. According to Gneiting and Raftery (2007), a good probabilistic forecasting system should be reliable and sharp. Reliability refers to the statistical consistency between the predictive distribution and the observation, while sharpness refers to the concentration of the predictive distribution. A reliable probabilistic forecasting system produces predictive distributions which are unbiased and representative of the true uncertainty underlying the process. These two attributes are important in an operational context, as scenarios are used for decision making. In the following, the term “dispersion” refers to the spread of the ensemble forecasts.

Forecast reliability is assessed using the reliability diagram. The reliability diagram diagnostic tool compares the observed coverage frequency (effective,  $1 - \hat{\alpha}$ ) with the corresponding theoretical confidence levels (nominal,  $1 - \alpha$ ) of predictive confidence intervals calculated from ensemble forecasts. Of course, if forecasts are reliable, the values  $1 - \hat{\alpha}$  and  $1 - \alpha$  should be equal for any confidence level. Moreover, the probability integral transform (PIT) histogram, which has the same interpretation as the rank histogram described in Hamill (2001), is also used to detect bias and dispersion issues in forecasts.

Scoring rules address reliability and sharpness simultaneously. One of the most well-known probabilistic scoring rules, the CRPS (Matheson and Winkler, 1976) is used to assess the overall accuracy of competing forecasting systems. The mathematical expression for the CRPS is given by Eq. (1).

$$\text{CRPS}(p(x), y) = \int (p(x) - H(x < y))^2 dx, \quad (1)$$

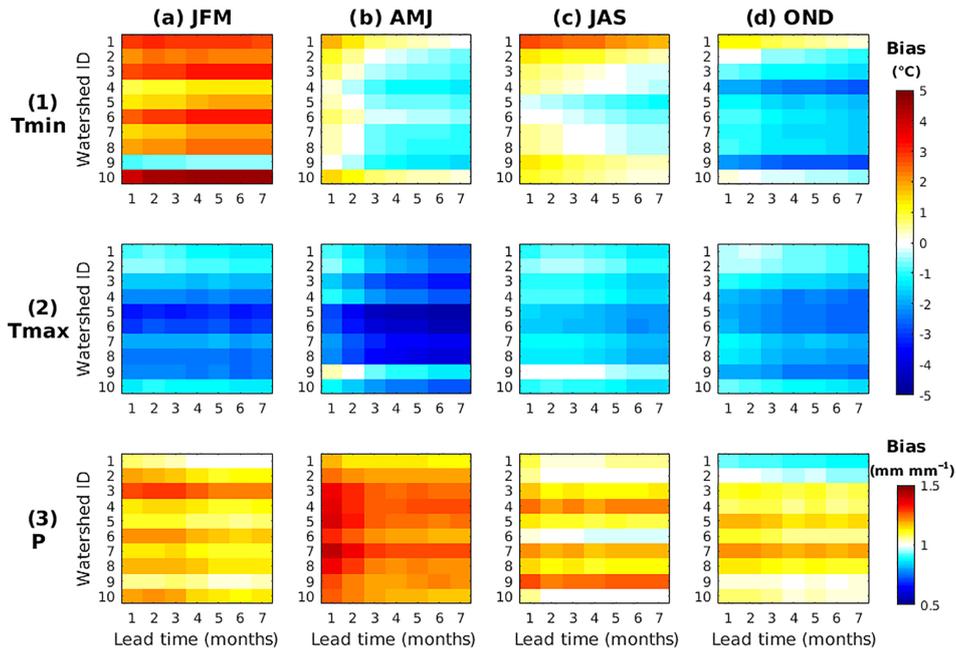
where  $p(x)$  represents the cumulative predictive distribution of the forecast and  $y$  is the observation.  $H$  is the step function, which equals 0 when  $x < y$  and 1 when  $x > y$ .

The continuous ranked probability skill score (CRPSS) expression is expressed as follows:

$$\text{CRPSS} = 1 - \frac{\text{CRPS}_{\text{for}}}{\text{CRPS}_{\text{ref}}}, \quad (2)$$

where  $\text{CRPS}_{\text{for}}$  is the mean CRPS of the forecasting system and  $\text{CRPS}_{\text{ref}}$  is the mean CRPS of the reference system (benchmark).

In order to evaluate the potential of ensemble meteorological forecasts, simulated streamflows were used instead of observations in the verification process. Proceeding in this way eliminates concerns about model and parametrization errors, which vary with watershed and time of the year. Moreover, after a spin-up period, the initial conditions are not necessarily estimated correctly by the hydrological model. Operationally, this is corrected by the forecaster, manually or using an automated data assimilation procedure, so that the simulation matches the observations closely. Since data assimilation falls outside the scope of the present study, using



**Figure 2.** Bias for forecasts of (1) monthly mean minimal temperature, (2) monthly mean maximal temperature and (3) monthly accumulated precipitation for the 10 watersheds under study as a function of seasons and lead times.

simulated streamflow as a benchmark eliminates this concern. Furthermore, as it is frequently the case for hydropower reservoirs, observations are not really obtained from gauging stations but rather estimated by a water balance equation applied on each reservoir. The quality of observations also varies across watersheds. Simulated streamflow series form a complete data set with no missing data but are subject to errors attributable in a large portion to the hydrological model itself. Consequently, the results presented in Sect. 5 should be interpreted as the potential skill (not the operational one) of meteorological forcings, as if the hydrological model was able to reproduce the watershed’s behaviour perfectly.

In the following, raw and bias-corrected meteorological forecasts will refer to the ensemble meteorological forecasts of the ECMWF System 4. Hydrological forecasts based on bias-corrected forecasts from System 4 will be referred to as corr-DSP. Hydrological forecasts based on climatology will be referred to as ESPs, and hydrological forecasts based on the simulated streamflow climatology as simulated HSP or sim-HSP. The lead time is defined herein as the time between the date of emission of the forecast and the end of the validity period of the forecasts. For instance, the 1-month lead time of the forecast issued on 1 January is the monthly volume or temperature of January.

## 5 Results

### 5.1 Bias characterisation and correction

Raw forecasts from ECMWF System 4 suffer from biases. Bias is calculated as shown in Eq. (3) for temperature and Eq. (4) for precipitation.

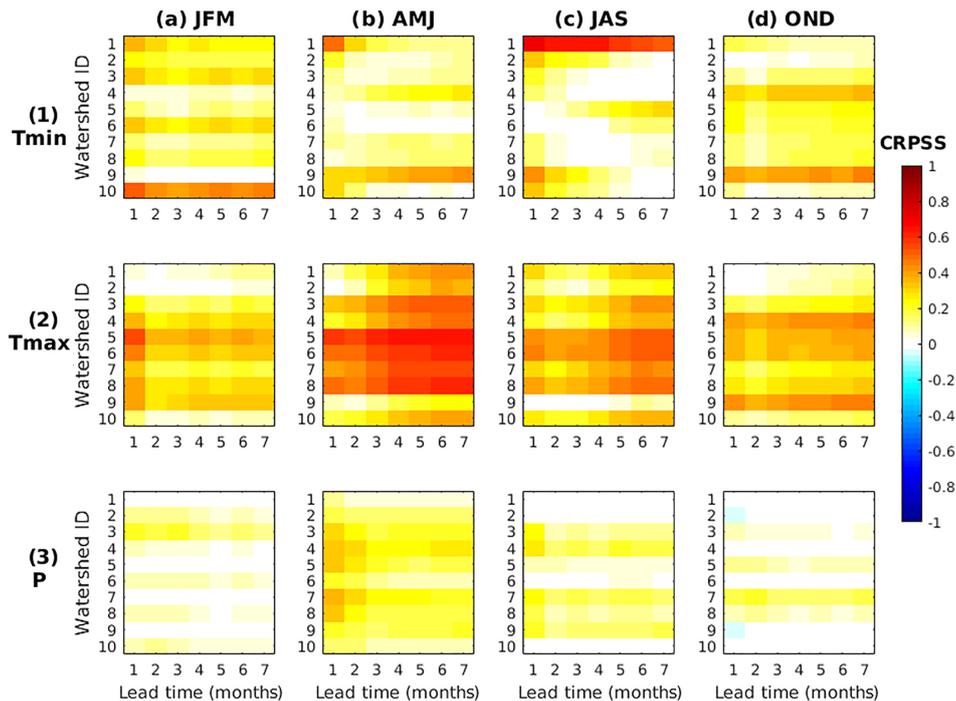
$$\text{Bias}_{\text{Add}} = \frac{1}{N} \sum_{k=1}^N (\bar{x}(k) - y(k)), \tag{3}$$

$$\text{Bias}_{\text{mult}} = \frac{\frac{1}{N} \sum_{k=1}^N \bar{x}(k)}{\frac{1}{N} \sum_{k=1}^N y(k)}, \tag{4}$$

where  $k$  is the day index,  $\bar{x}(k)$  the mean of the ensemble forecast,  $y(k)$  the observation and  $N$  the number of forecast–observation pairs considered for the bias estimation. We therefore assume an additive bias for monthly mean minimal and maximal temperature, the mean error. For monthly precipitation, bias is defined as the ratio of the forecasts mean to the mean observed accumulation. A multiplicative bias is then assumed for this meteorological variable.

Figure 2 shows the bias for monthly forecasts by season, lead time and watershed.

Monthly mean maximal temperature forecasts exhibit a cold bias for all watersheds, seasons and lead times. This cold bias increases with the lead time. Bias for the mean minimal temperatures changes depending on the season. For almost all watersheds, raw forecasts display a warm bias dur-



**Figure 3.** CRPSS of bias-corrected System 4 forecasts compared to raw ensemble forecasts of (1) monthly mean minimum temperature, (2) monthly mean maximum temperature and (3) monthly cumulative precipitation by watershed, season and lead time.

ing winter and a cold bias during fall. During spring and summer, temperature forecasts are almost unbiased. Monthly accumulated precipitation are overestimated most of the time, especially during spring.

As shown in Fig. 2, raw forecasts clearly need to be bias corrected. As mentioned above, Crochemore et al. (2016) have shown that the simple linear scaling method provides results comparable to the more complex distribution mapping to correct the bias in precipitation ensemble forecasts. In this study, daily precipitation forecasts as well as minimal and maximal temperatures are corrected by linear scaling based on monthly bias. Bias is estimated separately for each lead time and month of the year. A leave-one-year-out procedure is used to calculate bias and correct the forecasts. This consists of calculating the bias based on available forecasts issued on the same month, excluding the month under correction. For instance, for a given forecast, all other forecasts issued on the same day of the year are used to quantify the bias, calculating the mean of the errors between the ensemble mean and the observation. Multiplicative bias for precipitation and additive bias for temperature are calculated for each monthly forecast. The computed bias is finally used to correct the original daily forecasts.

Figure 3 presents the CRPSS of bias-corrected forecasts. The raw ensemble forecasts is taken as the reference (see Eq. 2). A CRPSS above 0 (from yellow to red) indicates that bias correction improves the original forecasts, whereas

a CRPSS below 0 (from light blue to dark blue) indicates a deterioration of the forecasts.

Bias correction is found effective, as it does improve meteorological forecasts according to the CRPSS. The effect of linear scaling for monthly forecasts is not homogeneous throughout the year and it also varies among watersheds. Bias correction is particularly efficient for periods with substantial bias, such as monthly aggregated precipitation during spring. Bias correction of monthly maximum temperature is also efficient for all months, watersheds and lead times. However, for precipitation during winter and fall, bias correction does not improve the CRPS noticeably. This is likely because biases during those months are generally small.

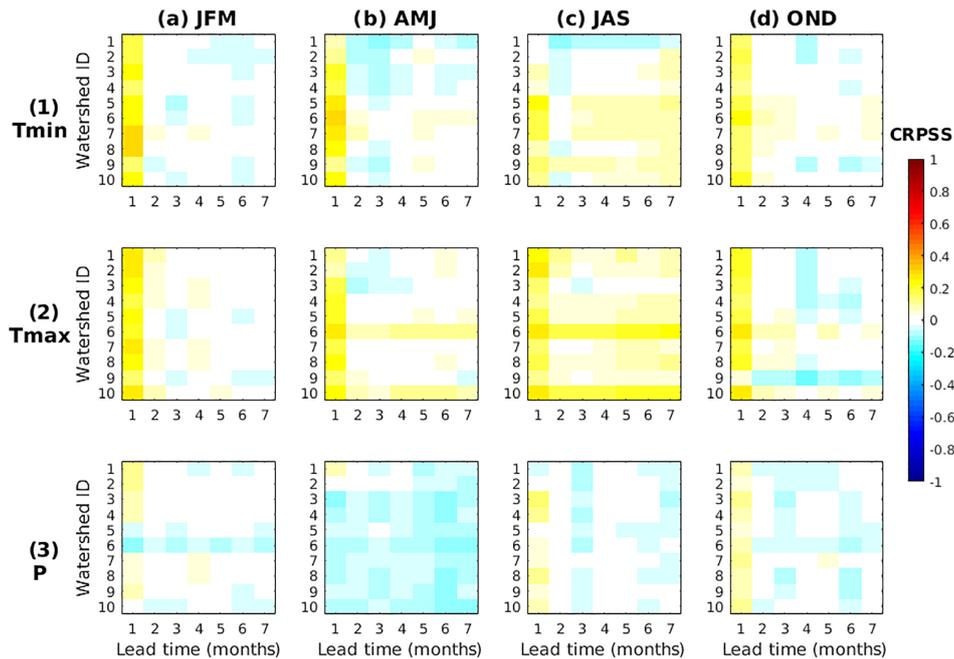
## 5.2 Performance of ensemble forecasts

### 5.2.1 Bias-corrected meteorological ensemble forecasts compared against climatology

Figure 4 presents the CRPSS of bias-corrected ensemble forecasts with climatology taken as the reference.

For almost all watersheds and seasons, both mean minimum and maximum temperatures outperform climatology for the 1-month lead time. However, for longer lead times, only temperature forecasts during summer can provide (only minor) improvement over climatology.

Precipitation is known to be less predictable than temperature and the CRPSS confirms this insight. For the 1-month lead time, CRPSS results are mixed. Ensemble fore-



**Figure 4.** CRPSS of bias-corrected System 4 forecasts compared to climatology for (1) monthly mean minimum temperature, (2) monthly mean maximum temperature and (3) monthly cumulative precipitation by watershed, season and lead time.

casts do have some skill for certain watersheds during winter, summer and fall (for instance watersheds number 3 and 4), whereas CRPSSs indicate that climatology is more skillful during spring. For longer lead times, according to the CRPSS, climatology always outperforms forecasts.

Hydrological conditions depend mostly on precipitation. However, in a northern environment, temperatures are also important, especially during winter and spring. In fact, during these periods, temperature defines the type of precipitation (snow or rain). It also drives snow pack evolution and the characteristics of the spring freshet (early/late and fast/slow). During summer, temperature controls evapotranspiration. In these conditions, it is valuable to assess the performance of hydrological forecasts produced by both systems: corr-DSP (based on bias-corrected System 4 forecasts) and ESP (based on climatology).

### 5.2.2 Monthly inflow volume forecasts

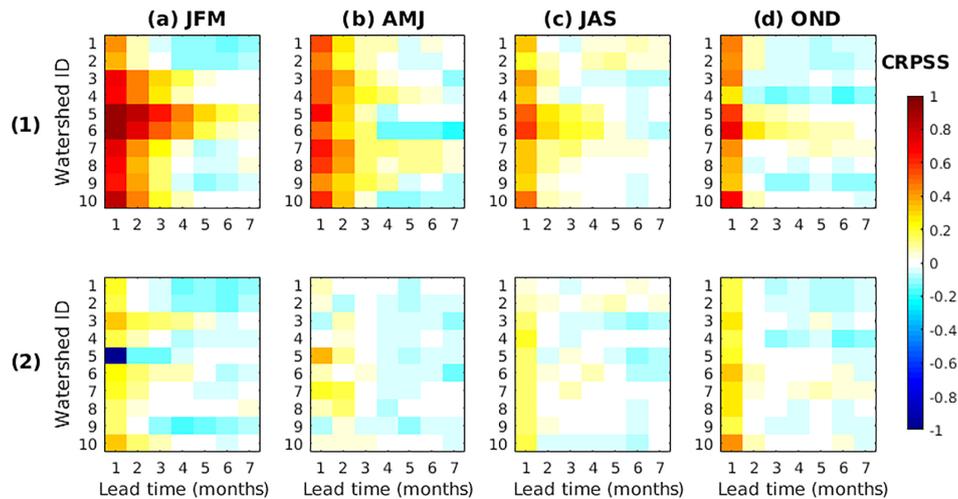
Figure 5 shows the performance of corr-DSP for inflow volumes compared to sim-HSP (simulated streamflow climatology) and ESP. Their performance is assessed using the CRPSS.

The first row of Fig. 5 shows the improvement of corr-DSP over sim-HSP. Those results reflect the gain in performance that could be achieved by considering initial conditions as well as information about bias-corrected meteorological forecasts from System 4. For all seasons and watersheds, it is valuable to use meteorological ensemble forecasts to produce monthly inflow volume forecasts instead of us-

ing simulated climatology (sim-HSP). The critical lead time, namely the lead time beyond which sim-HSP performs better than corr-DSP, depends on the period of the year. More specifically, inflow volume forecasts for summer and fall do not show much skill beyond the 1-month lead time. However, forecasts for winter and spring can be predicted fairly accurately several months ahead when using System 4 rather than sim-HSP.

When comparing corr-DSP with ESP (second row of Fig. 5), the CRPSS reflects the advantage of integrating meteorological information from ensemble forecasts into the hydrological model. The benefit is clear for the 1-month lead time, except during spring. The reasons why the CRPSS is not as good during spring as for other seasons could include a change in the influence of initial conditions during the different seasons and the lack of skill of precipitation forecasts. Indeed, streamflow is more variable during spring than during the rest of the year.

In general, corr-DSP outperforms ESP for the 1-month lead time with some exceptions, such as watershed number 5 in winter or watersheds number 3 and 9 during spring. Predicting monthly volume during summer and fall more than 1 month in advance is difficult and both ESP and corr-DSP exhibit comparable skill. Finally, for some watersheds during the winter months, corr-DSP improves the predictability of monthly volumes compared to ESP. Watersheds 3, 5 and 7 reflect different CRPSS behaviours, especially for winter and spring months. Thus, for the remainder of the analysis, special attention is given to those three specific watersheds.



**Figure 5.** CRPSS of ensemble forecasts of monthly inflow volume for reservoirs, produced by corr-DSP compared to (1) sim-HSP and (2) ESP. CRPSSs are shown by watershed, season and lead time.

Figure 6 presents the reliability diagrams of corr-DSP for three specific watersheds (numbers 3, 5 and 7; see Fig. 1) by season and lead time. The same diagrams were plotted for all 10 watersheds, but results are shown only for those three because, as mentioned above, they reflect specific behaviours worthy of investigation. Furthermore, in retrospect it was found that reliability diagrams for the other watersheds displayed characteristics quite similar to those that are presented. Reliability plots were also obtained for sim-HSP and ESP but they are, by definition, reliable. Hence, those plots are not shown.

As can be seen in Fig. 6, the reliability of corr-DSP monthly inflow volume forecasts changes over the seasons. The forecasts produced for the fall season are the most reliable, as the effective probabilities computed from forecasts are close to nominal probabilities for all lead times. Forecasts do not display strong dispersion issues. However, small differences between the observed coverage frequency and the theoretical confidence level remain. Indeed, even if bias correction was applied to raw meteorological forecasts, some biases can still remain and propagate to hydrological forecasts. In addition, bias correction affects the dispersion of precipitation forecasts and, in turn, the dispersion of hydrological forecasts. Figure 7 shows the PIT histograms for the three watersheds for the 1-month lead time. A flat PIT histogram corresponds to an accurate forecasting system whereas a higher effective frequency on one side of the histogram indicates the presence of bias (asymmetric shape). Higher effective frequency in the middle of the PIT histogram is linked with too much dispersion of the ensemble (bell-shaped) and, conversely, higher effective frequencies on both sides of the histogram are the sign of an under-dispersive behaviour (U-shaped; the spread of the ensemble is too small for most forecasts). As mentioned above, some biases are visible in

some cases, such as for watershed number 3 during the winter (Fig. 7, panel 1a) or under-dispersive behaviour during the spring (Fig. 7, panel 1b).

### 5.2.3 The case of inflow volume forecasting during spring freshet: an example for watersheds number 3, 5 and 7

Anticipating the inflow volume to the reservoirs is crucial for winter dam management. During the winter, reservoirs are partly emptied to ensure storage space for the inflow volume that is expected during the spring freshet. Consequently, good forecasts of inflow volumes are valuable. The spring freshet volume is computed by cumulating daily forecasted volumes over 3-month periods. The specific time period associated with the spring freshet varies from one watershed to another, mainly because of geographical location. Watersheds have been clustered into two groups. Watersheds 1 and 2, located in the south, have earlier spring freshet. For those two watersheds, the spring freshet period is defined from 1 March to the end of May. For the other watersheds, the spring freshet occurs between 1 April and the end of June.

Figures 8, 9 and 10 present the box plots of ensemble forecasts for the spring freshet volume from (a) ESP and (b) corr-DSP for three watersheds (numbers 3, 5 and 7). In those figures, the 3-month lead-time forecast corresponds to the forecasts issued at the beginning of the spring freshet period with a 3-month validity period. The 4- and 5-month lead-time forecasts are issued, respectively, 1 and 2 months before the beginning of the spring freshet period, with corresponding 4- and 5-month validity period. For ESP and corr-DSP, the dispersion of forecasts increase with the lead time. For those three watersheds, ESP exhibits a larger dispersion at all lead times. Extreme meteorological scenarios from past years lead to possible extreme hydrological scenarios. In

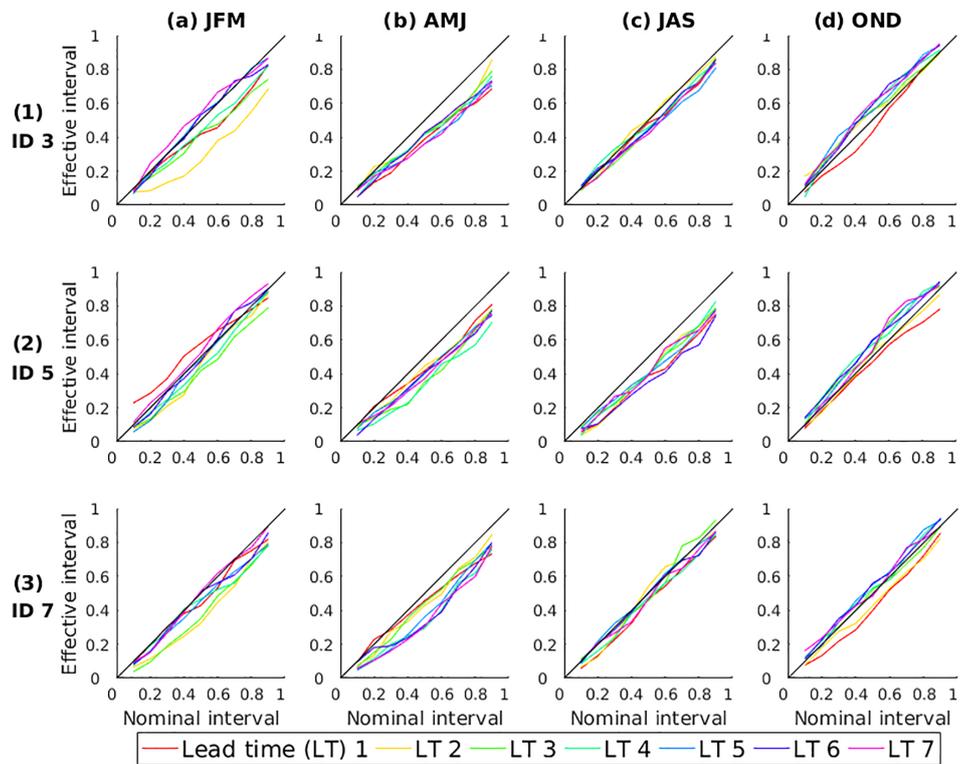


Figure 6. Reliability diagrams for (1) watershed number 3, (2) watershed number 5 and (3) watershed number 7.

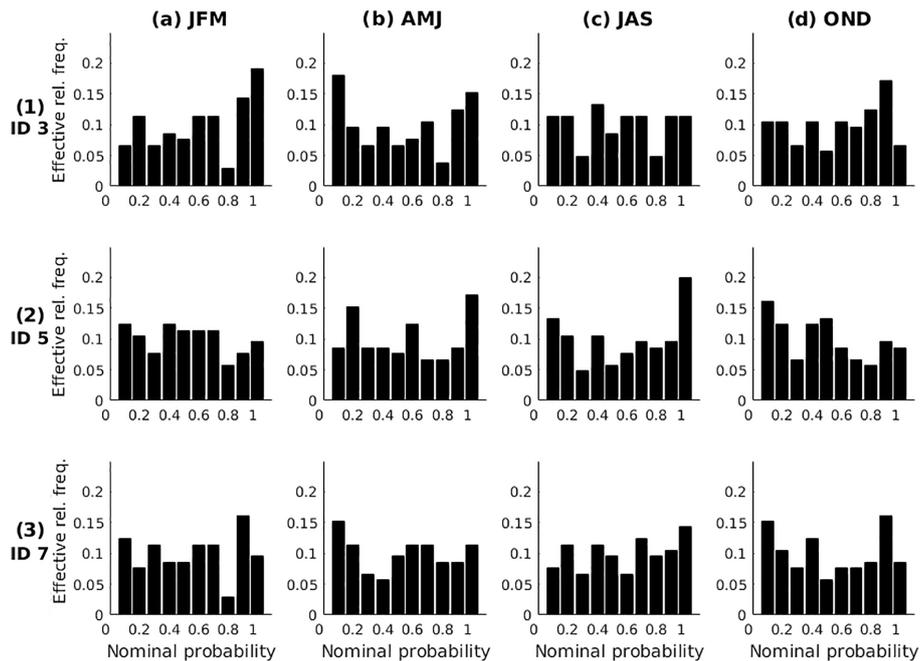
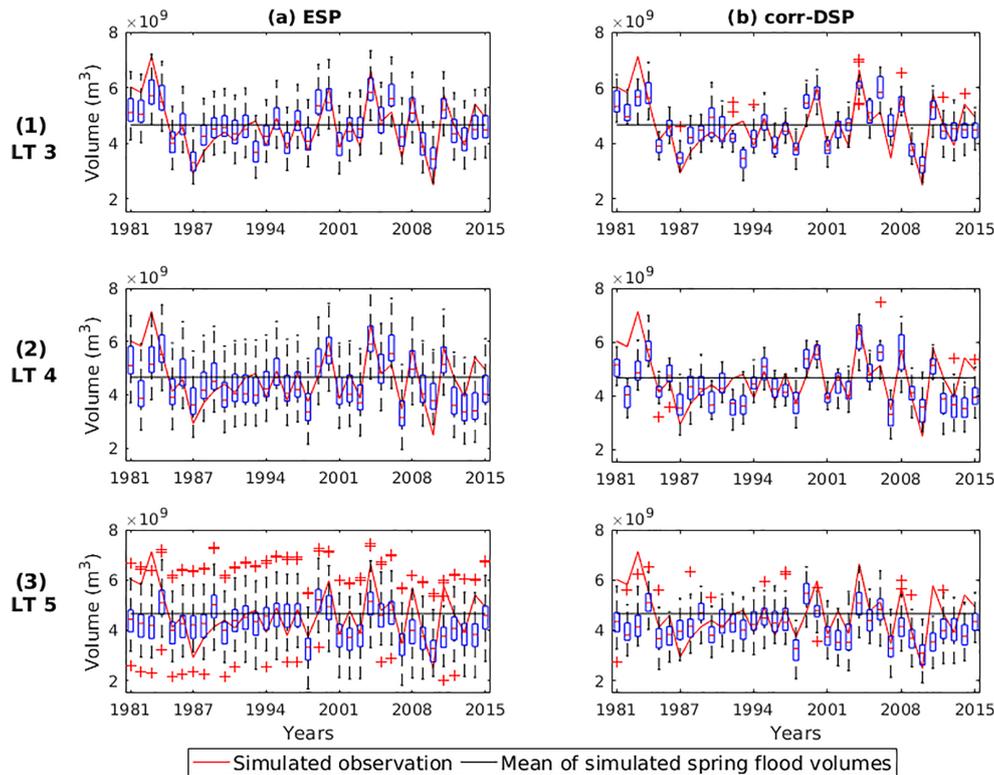


Figure 7. PIT histograms for (1) watershed number 3, (2) watershed number 5 and (3) watershed number 7.



**Figure 8.** Spring freshet volume forecasts at (1) 3-month lead time, (2) 4-month lead time and (3) 5-month lead time for watershed number 3. The box plots represent the ensemble forecasts for one given spring freshet, and the red line the corresponding simulated observations.

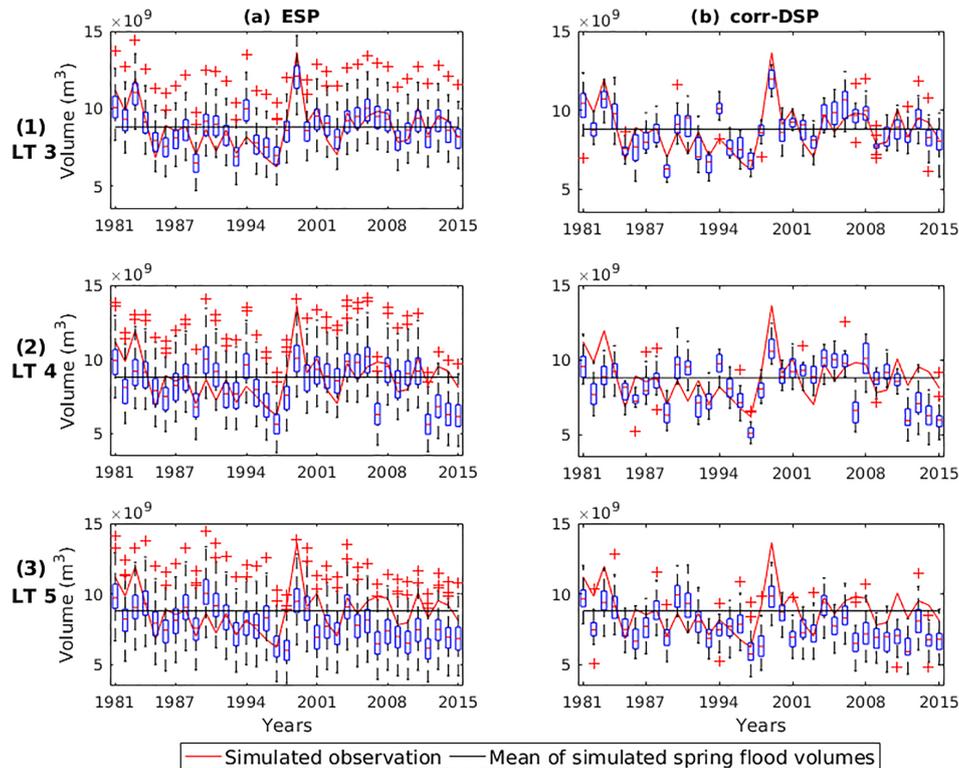
some cases, such as the spring of 1993 for watershed number 3 at lead time 3, corr-DSP provided very poor forecasts that missed the spring freshet almost entirely, whereas the ESP is much more successful (the observed volume is included in the box plot). This issue could be explained from a bias-correction problem. Indeed, even if bias correction was applied, some biases in corr-DSP might remain and further propagate to hydrological forecasts. It is also possible that the bias-correction method performs better for some years, lead times and watersheds than others. Moreover, single events at timescales smaller than 1 month might be subject to biases different than the monthly values used for bias correction. Similar figures were obtained for the other watersheds and the general conclusions for those figures are that corr-DSP exhibits a lower dispersion than ESP, leading observations to fall outside the boundary of the predictive distribution too often. As precipitation is overpredicted most of the time, as shown in Fig. 2, bias-corrected precipitation forecasts exhibit a lower dispersion than raw forecasts. This can explain the smaller dispersion of the volume forecasts.

Figure 11 presents the box plots of the CRPS for the 35 spring freshet events between 1981 and 2015, for all watersheds and the three forecasting systems: sim-HSP, ESP and corr-DSP. In addition to the forecasts issued the first day of the spring freshet period (3-month lead time), three lead times are considered: 1, 2 and 3 months before the first day of

the spring freshet. In most cases, the CRPS for corr-DSP displays a higher variability than the CRPS for ESP. Depending of the year, corr-DSP can have a better or worse performance than ESP and even sim-HSP. Hence, resorting to sim-HSP, the simplest “forecasting” system of all three (arguably not a forecasting system at all) can be advantageous for some watersheds for which the predictability is low. For instance, for watershed 6 at lead times 5 and 6 months, sim-HSP outperforms all other techniques. However, for lead times 3 and 4, ESP and corr-DSP have lower CRPSs in the majority of cases. Very high (poor) CRPSs can result from both dispersion and bias issues.

## 6 Conclusion

The objective of this study was to compare the performance and the behaviour of three hydrological forecasting systems for 10 watersheds in a northern climate (Québec, Canada). The three forecasting systems consist of HSP (streamflow climatology), ESP (forecasts based on meteorological climatology) and DSP (forecasts based on ensemble meteorological forecasts from ECMWF System 4). Streamflow simulations were used to build the streamflow climatology of each watershed. Simulated streamflows were also used as



**Figure 9.** Spring freshet volume forecasts at (1) 3-month lead time, (2) 4-month lead time and (3) 5-month lead time for watershed number 5. The box plots represent the ensemble forecasts for one given spring freshet, and the red line the corresponding simulated observations.

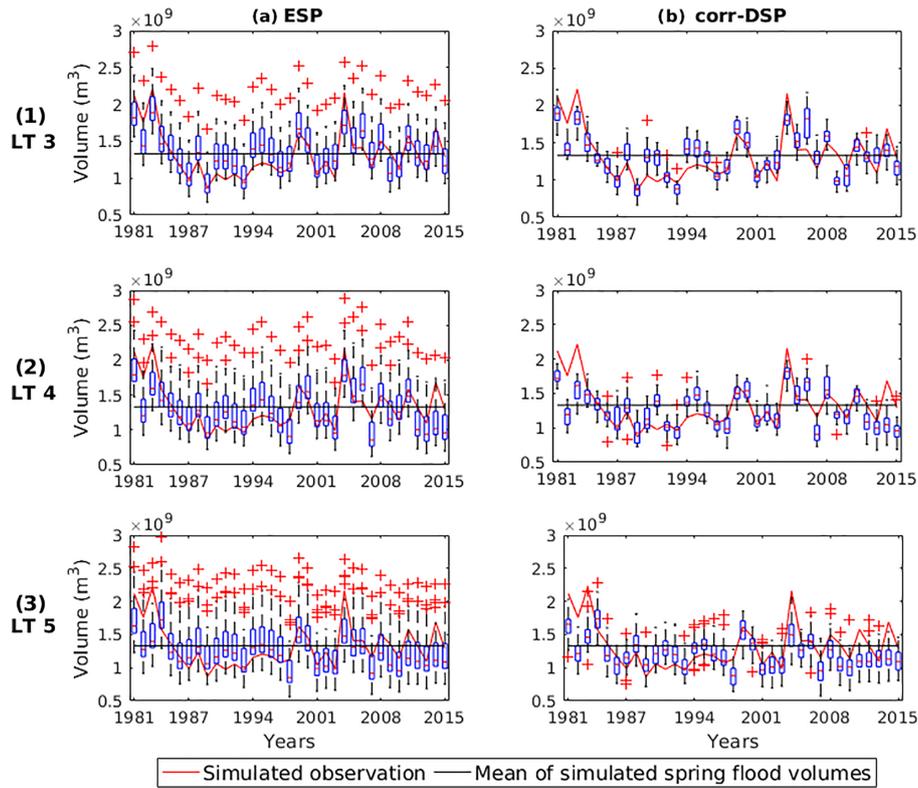
pseudo observations in order to avoid considering hydrological model errors in the analysis.

In the context of this study, it was found that ensemble meteorological forecasts from System 4 suffer from biases (see Fig. 2). However, bias correction, performed using the linear scaling method, results in an improved performance of ensemble meteorological forecasts, as assessed by the CRPS (see Sect. 4). Monthly mean minimal and maximal temperature forecasts outperform climatology for the 1-month lead time. The predictability extends to several months in specific cases (watersheds and seasons). Monthly accumulated precipitation is less predictable. In fact, ensemble meteorological forecasts do not have skill when it comes to forecasting monthly precipitation during spring. For other seasons, they slightly outperform climatology for the 1-month lead time.

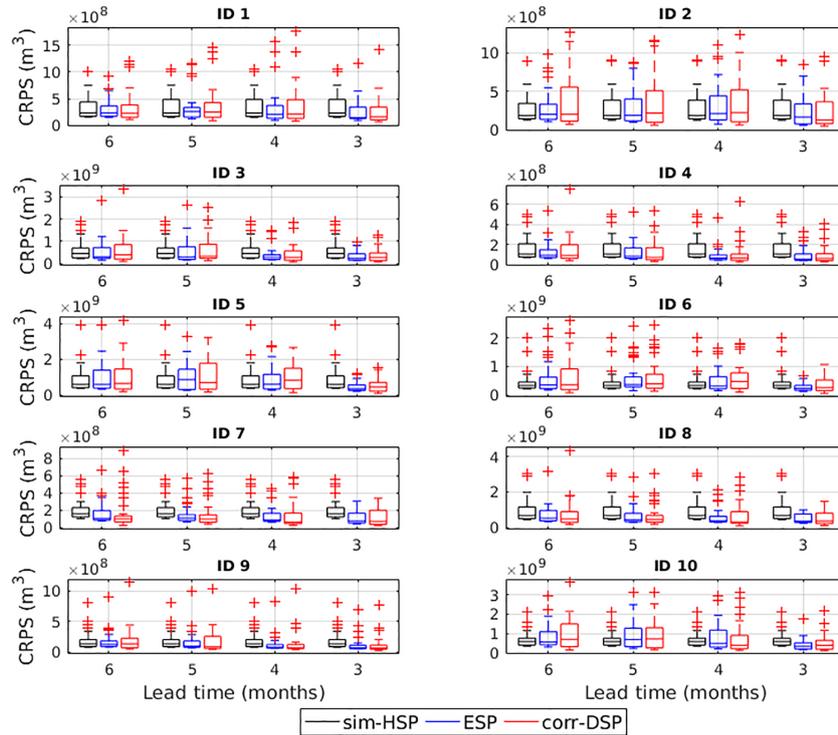
Still, according to the CRPS, bias-corrected ensemble meteorological forecasts were found to be a useful source of information to improve monthly volume forecasts (see Fig. 5), especially for the 1-month lead time. This is likely due to temperature forecasts more than precipitation forecasts, as mentioned above. Regarding the particular case of forecasts for summer and fall, the CRPS of corr-DSP outperforms the CRPS of both sim-HSP and ESP for the 1-month lead time. Predictability of monthly flow volume for winter and spring months extends up to 3 months against simulated climatology. The CRPS between corr-DSP and ESP is lower

than the CRPS between corr-DSP and sim-HSP. However, corr-DSP shows some skill from 1-month lead time up to 2- or 3-month lead times for some watersheds (numbers 3, 7 and 10) during the winter. Monthly forecasts based on System 4 are less reliable than ESP, and this possibly originates from bias propagation or dispersion issues (see Fig. 7 for an example) in both meteorological and hydrological forecasts. The lack of skill of corr-DSP can originate from different sources. First, linear scaling is a rather simple bias-correction method. It was performed using monthly bias and thus there is a possibility that biases at smaller temporal scales could remain. Second, as precipitation was originally over-predicted in most cases by System 4 (see Fig. 2), bias-correction results in a reduction of the ensemble spread for precipitation forecasts, and possibly for streamflow forecasts as well. Results for spring are mixed: the forecasting performance during spring freshet varies from one watershed to another. In general, the CRPS of corr-DSP is more variable than the CRPS of ESP.

Furthermore, skill scores are subject to sample uncertainty. In this study, each skill score computation is based on almost 100 forecasts. The size of the verification set thus remains a limit for assessing the significance of the verification results. A bootstrap procedure was performed to assess the significance of the difference in CRPS between corr-DSP and ESP. Results show that in most cases, results are not proven



**Figure 10.** Spring freshet volume forecasts at (1) 3-month lead time, (2) 4-month lead time and (3) 5-month lead time for watershed number 7. The box plots represent the ensemble forecasts for one given spring freshet, and the red line the corresponding simulated observations.



**Figure 11.** Box plots of the CRPS for the 35 spring freshet events between 1981 and 2015 for all watersheds and the three forecasting systems: sim-HSP (black), ESP (blue) and corr-DSP (red).

to be significant and confirm the previous insight. For the 1-month lead time, the quality of corr-DSP is clearer and for longer lead times, the CRPSS of corr-DSP compared to ESP tends to 0. Consequently, in this study and for long lead times, it is not clear that bias-corrected seasonal meteorological ensemble forecasts from a dynamical model (System 4) can completely replace ESP. However, they provide substantial complementary information to produce long-term hydrological forecasts, as shown by the special case of forecasting the inflow volume associated with the spring freshet (see Sect. 5.2.3). An analysis on the economic value of forecasts for hydropower production, for instance using a reservoir operation model based on stochastic dynamic programming, would be the next logical step. It would allow one to determine whether or not the differences observed here between the three concurrent forecasting systems are indeed notable for water management.

Finally, ECMWF System 4 does not include any sea-ice model (ECMWF, 2017). This could limit improvement in weather predictability in mid-latitudes. Other providers of ensemble meteorological forecasts exist and a multi-model approach could improve the skill of the seasonal forecasts. Moreover, according to our results, it is for the 1-month lead time that the most important gain could be achieved by using dynamical models over climatology (ESP). For that purpose, the newly available Sub-seasonal to Seasonal (S2S) database (Vitart et al., 2017), which gathers ensemble forecasts from different agencies at the sub-seasonal scale (from 1 up to 60 days), could be explored. In fact, the forecasts available in S2S are especially tailored for the 1- to 2-month lead times, and hence could have superior skill for hydrological applications than System 4, which, as shown here, can often lead to better 1-month-ahead streamflow and volume forecasts than ESP. This new database would also enable future studies to explore multi-model forecasting approaches at long lead times and assess the ability of such an approach to extend the limits of predictability. Finally, as proposed by Yuan et al. (2014), forecasts for different lead times would need to be efficiently joined together in a seamless way, and there is much to explore in this regard.

*Data availability.* Unfortunately, the data used in this study are not publicly available. Data from the ECMWF System 4 forecasts are produced by the ECMWF but are not included in the repositories of public data sets. They were provided for the purpose of this study by Florian Pappenberger (florian.pappenberger@ecmwf.int). Meteorological and streamflow data as well as the watershed delineation file and the hydrological model are the property of Hydro-Québec.

*Author contributions.* RB performed all the computations, prepared all figures and wrote the most of the manuscript. MAB guided the work, provided opinions about the presentation of results and analysis to be done as well as codes that were used in the computations and helped write the manuscript and revised all versions.

LP provided codes and further guidance for analysing the results and preparing the manuscript. RL initiated the work, reviewed the manuscript and participated in analysing the results.

*Competing interests.* The authors declare that they have no conflict of interest.

*Special issue statement.* This article is part of the special issue “Sub-seasonal to seasonal hydrological forecasting”. It is a result of the HEPEX workshop on seasonal hydrological forecasting, Norrköping, Sweden, 21–23 September 2015.

*Acknowledgements.* The authors want to thank Kristian Förster and two anonymous reviewers for their useful comments on this article. This work was funded by a NSERC Cooperative Research and Development grant to Robert Leconte. Rachel Bazile gratefully acknowledges a scholarship from the Fonds de Recherche du Québec Nature et Technologies. The authors wish to thank Florian Pappenberger for providing the forecasts from ECMWF System 4 as well as Catherine Guay for her support with HSAMI. The team of operational forecasters at Hydro-Québec are warmly acknowledged for their insights, especially Fabian-Tito Arandia Martinez, Éric Crobeddu and Marie-Claude Simard.

Edited by: Quan J. Wang

Reviewed by: Kristian Förster and two anonymous referees

## References

- Bröcker, J. and Smith, L. A.: From ensemble forecasts to predictive distribution functions, *Tellus A*, 60, 663–678, 2008.
- Cloke, H. and Pappenberger, F.: Ensemble flood forecasting: a review, *J. Hydrol.*, 375, 613–626, 2009.
- Crochemore, L., Ramos, M.-H., and Pappenberger, F.: Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts, *Hydrol. Earth Syst. Sci.*, 20, 3601–3618, <https://doi.org/10.5194/hess-20-3601-2016>, 2016.
- Day, G. N.: Extended streamflow forecasting using NWSRFS, *J. Water Res. Plan. Man.*, 111, 157–170, 1985.
- DelSole, T.: Predictability and information theory. Part I: Measures of predictability, *J. Atmos. Sci.*, 61, 2425–2440, 2004.
- Demargne, J., Wu, L., Regonda, S. K., Brown, J. D., Lee, H., He, M., Seo, D.-J., Hartman, R., Herr, H. D., Fresch, M., et al.: The science of NOAA’s operational hydrologic ensemble forecast service, *B. Am. Meteorol. Soc.*, 95, 79–98, 2014.
- ECMWF: available at: [https://www.ecmwf.int/sites/default/files/medialibrary/2017-10/System4\\_guide.pdf](https://www.ecmwf.int/sites/default/files/medialibrary/2017-10/System4_guide.pdf), last access: 15 November 2017.
- Fortin, V.: Le modèle météo-apport HSAMI: historique, théorie et application, Institut de recherche d’Hydro-Québec, Varennes, 2000.
- García-Morales, M. B. and Dubus, L.: Forecasting precipitation for hydroelectric power management: how to exploit GCM’s seasonal ensemble forecasts, *Int. J. Climatol.*, 27, 1691–1705, 2007.

- Gneiting, T. and Raftery, A. E.: Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.*, 102, 359–378, 2007.
- Guay, C., Minville, M., and Braun, M.: A global portrait of hydrological changes at the 2050 horizon for the province of Québec, *Can. Water Resour. J.*, 40, 285–302, 2015.
- Hamill, T. M.: Interpretation of rank histograms for verifying ensemble forecasts, *Mon. Weather Rev.*, 129, 550–560, 2001.
- He, M., Whitin, B., Hartman, R., Henkel, A., Fickenschers, P., Staggs, S., Morin, A., Imgarten, M., Haynes, A., and Russo, M.: Verification of Ensemble Water Supply Forecasts for Sierra Nevada Watersheds, *Hydrology*, 3, 35, 2016.
- Hydro-Québec: Rapport annuel, available at: <http://www.hydroquebec.com/publications/fr/documents-entreprise/rapport-annuel.html> (last access: 5 September 2017), 2015.
- Kim, H.-M., Webster, P. J., and Curry, J. A.: Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter, *Clim. Dynam.*, 39, 2957–2973, 2012.
- Lepage, M.-P. and Bourgeois, G.: Le réseau québécois de stations météorologiques, Centre de référence en agriculture et agroalimentaire du Québec (CRAAQ), p. 15, available at: <https://www.craaq.qc.ca/data/DOCUMENTS/PAGR0101.pdf> (last access: 21 November 2017), 2011.
- Luo, L. and Wood, E. F.: Use of Bayesian merging techniques in a multimodel seasonal hydrologic ensemble prediction system for the eastern United States, *J. Hydrometeorol.*, 9, 866–884, 2008.
- Marty, R., Zin, I., Obled, C., Bontron, G., and Djerboua, A.: Toward real-time daily PQQF by an analog sorting approach: application to flash-flood catchments, *J. Appl. Meteorol. Clim.*, 51, 505–520, 2012.
- Matheson, J. E. and Winkler, R. L.: Scoring rules for continuous probability distributions, *Manage. Sci.*, 22, 1087–1096, 1976.
- Molteni, F., Stockdale, T., Balmaseda, M., Balsamo, G., Buizza, R., Ferranti, L., Magnusson, L., Mogensen, K., Palmer, T., and Vitart, F.: The new ECMWF seasonal forecast system (System 4), European Centre for Medium-Range Weather Forecasts, 2011.
- Moradkhani, H. and Meier, M.: Long-lead water supply forecast using large-scale climate predictors and independent component analysis, *J. Hydrol. Eng.*, 15, 744–762, 2010.
- Ouranos: Vers l'adaptation. Synthèse des connaissances sur les changements climatiques au Québec. Partie 1 : Évolution climatique au Québec., p. 114, 2015.
- Palmer, T. N., Doblus-Reyes, F. J., Hagedorn, R., Alessandri, A., Gualdi, S., Andersen, U., Feddersen, H., Cantelaube, P., Terres, J., Davey, M., Graham, R., Délecluse, P., Lazar, A., Déqué, M., Guérémy, J., Díez, E., Orfila, B., Hoshen, M., Morse, A. P., Keenlyside, N., Latif, M., Maisonnave, E., Rogel, P., Marletto, V., and Thomson, M. C.: Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEME-TER), *B. Am. Meteorol. Soc.*, 85, 853–872, 2004.
- Shukla, S., Sheffield, J., Wood, E. F., and Lettenmaier, D. P.: On the sources of global land surface hydrologic predictability, *Hydrol. Earth Syst. Sci.*, 17, 2781–2796, <https://doi.org/10.5194/hess-17-2781-2013>, 2013.
- Singh, S. K.: Long-term Streamflow Forecasting Based on Ensemble Streamflow Prediction Technique: A Case Study in New Zealand, *Water Resour. Manag.*, 30, 2295–2309, 2016.
- Sveinsson, O. G., Lall, U., Fortin, V., Perrault, L., Gaudet, J., Zebiak, S., and Kushnir, Y.: Forecasting spring reservoir inflows in Churchill Falls basin in Quebec, Canada, *J. Hydrol. Eng.*, 13, 426–437, 2008.
- Svensson, C.: Seasonal river flow forecasts for the United Kingdom using persistence and historical analogues, *Hydrolog. Sci. J.*, 61, 19–35, 2016.
- Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., Déqué, M., Ferranti, L., Fucile, E., Fuentes, M., Hendon, H., Hodgson, J., Kang, H.-S., Kumar, A., Lin, H., Liu, G., Liu, X., Malguzzi, P., Mallas, I., Manoussakis, M., Mustrangelo, D., MacLachlan, C., McLean, P., Minami, A., Mladek, R., Nakazawa, T., Najm, S., Nie, Y., Rixen, M., Robertson, A. W., Ruti, P., Sun, C., Takaya, Y., Tolstykh, M., Venuti, F., Waliser, D., Woolnough, S., Wu, T., Won, D.-J., Xiao, H., Zaripov, R., and Zhang, L.: The Subseasonal to Seasonal (S2S) Prediction Project Database, *B. Am. Meteorol. Soc.*, 98, 163–173, <https://doi.org/10.1175/BAMS-D-16-0017.1>, 2017.
- Wang, Q. J., Robertson, D. E., and Chiew, F. H. S.: A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites, *Water Resour. Res.*, 45, W05407, <https://doi.org/10.1029/2008WR007355>, 2009.
- Weisheimer, A. and Palmer, T.: On the reliability of seasonal climate forecasts, *J. R. Soc. Interface*, 11, 96, <https://doi.org/10.1098/rsif.2013.1162>, 2014.
- Wood, A. W. and Lettenmaier, D. P.: An ensemble approach for attribution of hydrologic prediction uncertainty, *Geophys. Res. Lett.*, 35, L14401, <https://doi.org/10.1029/2008GL034648>, 2008.
- Xu, C.-y.: Modelling the effects of climate change on water resources in central Sweden, *Water Resour. Manag.*, 14, 177–189, 2000.
- Yang, L., Tian, F., Sun, Y., Yuan, X., and Hu, H.: Attribution of hydrologic forecast uncertainty within scalable forecast windows, *Hydrol. Earth Syst. Sci.*, 18, 775–786, <https://doi.org/10.5194/hess-18-775-2014>, 2014.
- Yossef, N. C., Winsemius, H., Weerts, A., Beek, R., and Bierkens, M. F.: Skill of a global seasonal streamflow forecasting system, relative roles of initial conditions and meteorological forcing, *Water Resour. Res.*, 49, 4687–4699, 2013.
- Yuan, X., Wood, E. F., Roundy, J. K., and Pan, M.: CFSv2-based seasonal hydroclimatic forecasts over the conterminous United States, *J. Climate*, 26, 4828–4847, 2013.
- Yuan, X., Wood, E. F., and Liang, M.: Integrating weather and climate prediction: Toward seamless hydrologic forecasting, *Geophys. Res. Lett.*, 41, 5891–5896, 2014.
- Yuan, X., Wood, E. F., and Ma, Z.: A review on climate-model-based seasonal hydrologic forecasting: physical understanding and system development, *Wiley Interdisciplinary Reviews: Water*, 2, 523–536, 2015.
- Yuan, X., Ma, F., Wang, L., Zheng, Z., Ma, Z., Ye, A., and Peng, S.: An experimental seasonal hydrological forecasting system over the Yellow River basin – Part 1: Understanding the role of initial hydrological conditions, *Hydrol. Earth Syst. Sci.*, 20, 2437–2451, <https://doi.org/10.5194/hess-20-2437-2016>, 2016.