

An efficient method to correct under-dispersion in ensemble streamflow prediction of inflow volumes for reservoir optimization

Richard Arsenault^{1,2*}, Marco Latraverse² and Thierry Duchesne¹

1. Département de mathématiques et de statistique, Université Laval, 1045, av. de la Médecine, Québec (Québec), G1V 0A6, Canada.
2. Rio Tinto - Aluminium division, 1954 Davis, Jonquière, G7S 4R5, QC, Canada,

* Corresponding author email: richard.arsenault@riotinto.com

This is the final accepted version of the paper before journal typesetting. The final publication is available at Springer via <http://dx.doi.org/10.1007/s11269-016-1425-4>

Paper citation is as follows:

Arsenault, R., Latraverse, M. and Duchesne, T. (2016) An efficient method to correct under-dispersion in ensemble streamflow prediction for seasonal volumetric forecasting. *Water Resources Management*, 30(12), 4363–4380, doi: 10.1007/s11269-016-1425-4

Running head: Hindcasting-mode Uncertainty Estimation method for correcting ESP inflow volume under-dispersion

Keywords: ESP; Under-dispersion; Reservoir optimization; Ensemble forecasts; hydrologic modelling

Acknowledgements

This project was funded in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) under Engage Grant (EG) 479534 2015. The authors thank three anonymous reviewers for their helpful comments which helped shape the paper into its current form.

Abstract

Ensemble streamflow prediction (ESP) has been widely used to gain insight on possible future inflows to hydropower reservoirs. However underestimation of climate, model structure and initial condition uncertainty often leads to under-dispersed ESP forecasts. In this paper, we present a novel approach called “*Hindcast-mode Uncertainty Estimation*” (HUE) to efficiently add variability in ESP forecasts to reduce their under-dispersion. The method was tested on a Canadian catchment used by Rio Tinto – Aluminium division to produce hydropower for their aluminium smelting plants. This project was focused on correcting long-term predictions of freshet runoff volumes to optimize drawdown volumes, with up to 6 months of lead time. It was found that by adding an error term to the hydrological model’s snow water equivalent (SWE) state variable at the time of forecast in hindcasting mode, the resulting simulation could be forced to perfectly reproduce the freshet runoff volume. This error term was computed for all years on record which enabled modeling of the error’s distribution. This distribution can then be sampled from to add noise to the model’s SWE at the start of a new ESP forecast. Results show that the current winter ESP forecasts are strongly under-dispersed for the freshet runoff volume estimation and that the proposed method is able to widen the ESPs to correct the under-dispersion problem. This was validated by using Talagrand diagrams which shifted from a U-shape (prior to HUE) to a uniform distribution (with HUE). The project objectives of correcting the ESP forecast’s under-dispersion in spring runoff estimations was thus attained with minimal effort, bypassing the need to perform more complex ensemble data assimilation techniques.

1. Introduction

Hydroelectric power plants often rely on reservoirs to increase head and improve efficiency of the generating stations by managing water levels according to predicted inflows. Due to the stochastic nature and large uncertainty of future inflows, reservoir management experts may use ensemble streamflow prediction (ESP) methods to get a probabilistic view of possible outcomes (Day 1985). To generate an ESP forecast, a hydrological model is allowed to run on a spin-up period and its state variables are adjusted so as to be as close as possible to the day's real conditions. Then, a number of possible meteorological series are fed to the model, each one resulting in a different hydrograph. Meteorological time series can be weather forecasts (short-term ESPs), resampling of historical observations (mid/long-term ESPs), a mix of both, or time-series generated with stochastic weather generators (Chen and Brissette 2015). The resulting plume is the ESP forecast and, in ideal conditions, the ESP members will represent the possible outcomes with the same probability as the real world system. These ESP forecasts are ultimately used to assess probable streamflow outcomes used in water management decision-making (Faber and Stedinger 2001).

In the case of reservoir-based hydropower, policy optimization tools such as stochastic dynamic programming (SDP) and Sampling SDP (SSDP) require the ensemble members as inputs and then return the water drawdown decision that will generate the maximum expected profit based on the possible streamflow scenarios (Côté and Leconte 2015). Each member of the ESP may be weighted equally or according to predefined rules or user experience (Najafi et al. 2012). In either case, the optimization algorithm relies on the hypothesis that ESP members display the correct amount of expected variability in the possible outcomes so that the probability of ensemble inflows reflects the real-world probabilities of occurrence. If the ensemble is biased,

then inevitably the water drawdown decision will be suboptimal because the members will not represent the actual realization probabilities and the decision will be taken on wrong assumptions (Hashino et al. 2007).

A common problem in generating ESP forecasts is that of under-dispersion (or “overconfidence”). In such cases, the forecast plume is too narrow - often missing what will be the actual streamflow outcome. Under-dispersion is a symptom resulting from the fact that the ESP forecasts do not reproduce the real-world uncertainties adequately, thus the lack of variability in the predicted hydrographs. Under-dispersion is more common than over-dispersion because modelling systems are often attempts at simplifying the real world system, reducing sources of uncertainty in the process (Boucher et al. 2015). This paper will therefore focus specifically on the under-dispersion problem. There have been many attempts at correcting under-dispersion in ESP forecasting using aggregate measures, most of which have been successful for their particular application (Zalachori et al. 2012, Boucher et al. 2011). However, specific requirements must often be met for the method to be effective: model residual normality and homoscedasticity (Sloughter et al. 2006), linear relationships between state and output variables (Vrugt et al. 2006) or large numbers of model simulations (Velasquez et al. 2011), to name a few. Current ESP forecast under-dispersion correction methods can be categorized in two main classes: post-processing and pre-processing of the ensemble forecasts.

1.1 Post-processing of ensemble forecasts

The most commonly used approach in correcting ensemble forecast bias and under-dispersion is using post-processing methods. In such cases, ensemble traces are statistically enhanced as to increase their variability. Amongst the most widely recognized post-processing techniques are ensemble dressing (Roulston and Smith 2003, Wang and Bishop 2005, Fortin et al. 2006, Bröcker

and Smith 2008, Pagano et al. 2013), regression methods (Gneiting et al. 2005, Van Schaeybroeck and Vannitsem 2011) and Bayesian model averaging methods (Raftery et al. 2005). Boucher et al. (2015) recently compared a few existing as well as new methods to correct under-dispersion and bias in short-term continuous streamflow forecasting. Generally, these approaches tend to analyse past prediction errors and reintroduce those errors into the ensemble members through different means. While ensemble post-processing methods were shown to improve ensemble reliability and correct biases present in the forecasts, they share one major drawback for management policy optimization: relationships between the hydrological model's state variables and the dressed ensembles (and other statistically transformed members) are lost. For the system in the current study, the SDP algorithm requires hydrological state variables along with the inflows, namely the snow water equivalent and soil moisture (Côté et al. 2011, Desreumaux et al. 2014). By post-processing the ensemble members, the state variable becomes uncorrelated to the inflows and cannot be used in the SDP. It thus becomes impossible to relate the model's initial conditions to the model outputs through the transition variables that are used in the optimization algorithm (Côté and Leconte 2015). Therefore ensemble post-processing methods are not adequate solutions for our objectives.

1.2 ESP Pre-processing

Another option for generating variability in streamflow forecasts is to pre-process the inputs to the model. By adding noise to the inputs such as climate time-series or initial conditions, variability is inevitably increased. Chen et al. (2014) showed that calibrating a weather generator on historic weather records allowed producing large sets of climate data, freeing the ESP from short historical record constraints. Their method was shown to improve ensemble under-dispersion upon traditional ESP methods for short lead-times. However the forecasts remain

under-dispersed to a certain extent. Furthermore, resorting to weather generators increases simulation time and complexity by adding supplementary processes. This may be a limiting factor in an operational context where forecasts must often be run in short timeframes.

The alternative to adding variability to weather forcing is to add randomness to the model's initial conditions through ensemble data assimilation (Liu and Gupta 2007, Liu et al. 2015). In snowmelt-dominated catchments, Snow Water Equivalent (SWE) is a major factor and can have considerable impact in volumetric seasonal forecasts (Clark and Hay 2004, Franz et al. 2008, Zeinivand and De Smedt 2009, Sproles et al. 2016). DeChant and Moradkhani (2011) used a data assimilation approach to combine the traditional ESP framework with a probabilistic estimation of snow observations based on in-situ observations (SNOTEL), which is then used to generate new initial conditions. In this manner, they were able to improve ESP for seasonal probabilistic volumetric forecasts. However their method is limited by the available snow observations, which are also subject to measurement uncertainty. Roulin and Vannitsem (2015) recently showed that during winter, errors in the initial conditions are more important than those of model structure, confirming that adding variability to the initial conditions in the model snowpack is likely to increase the ESP forecast dispersion. Ensemble data assimilation for initial condition updating is limited by the historic observations which may not reasonably describe the actual basin state.

In this paper, we propose a novel method to use the hydrological model's initial conditions as an aggregator of all the different sources of uncertainty during the ESP forecasts. As was the case in DeChant and Moradkhani (2011), this paper focuses on volumetric forecasts during the snowmelt period on a snowmelt-dominated catchment. However, instead of using observations

of snowpack water content in a data assimilation framework, we estimate the probability density function of snowpack depth errors by analysing discrepancies between past simulation and observed inflows. The proposed method, which we refer to as Hindcast-mode Uncertainty Estimation (HUE), allows adding an adequate amount of noise to the ensembles to make them representative of the historic simulation error while maintaining a link between the model's state variables and the model outputs as required by the SDP reservoir optimization algorithm. HUE is also not dependent on any conditions such as distribution normality or homoscedasticity.

The next section will detail the study area and data. Section 3 will show the detailed methods proposed in this paper. Section 4 presents the study's main results, followed by a discussion and final concluding remarks.

2. Study area and data

2.1 Study area

The study area is a sub-basin of the 75,000 km² Lac-St-Jean watershed in central Québec, Canada. Covering 45,261 km², it is the most important catchment (area-wise) in Rio Tinto – Aluminium division's hydropower system in the region. Average daily discharge is 850m³/s, with peak floods topping 6,000m³/s. Figure 1 shows the catchment elevation map as well as the location of weather stations on the study site.

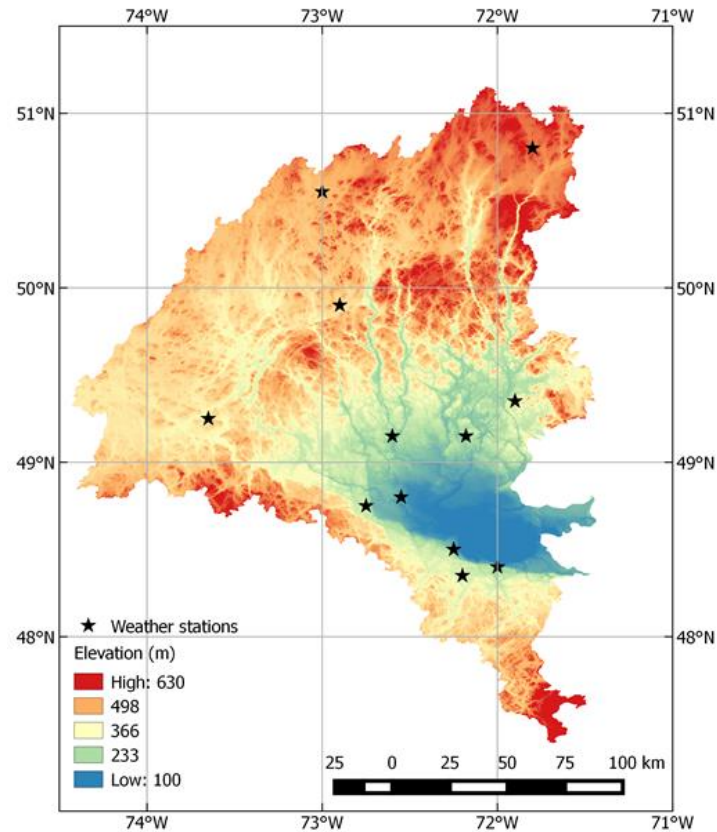


Fig. 1 Sub-catchment of the larger Lac-St-Jean watershed and location of weather stations

The catchments' northern location implies that the hydrological regime is snowmelt-dominated, which has a strong influence on water level management in the reservoirs (Poulin et al. 2011, Troin et al. 2015). Indeed, the freshet (spring melt) period can bring large volumes of water to the reservoir which must be emptied to a certain degree to accommodate the large inflows. Too large of a drawdown will have repercussions on hydropower production levels for the following months as the reservoir will not be at its optimal capacity, while too little of a drawdown might require wastefully spilling the inflows and could cause structural problems to the hydropower system (Arsenault et al. 2013). Therefore knowing freshet inflows is a crucial part of the long-term planning of reservoir operations.

2.2 Data

All data for this project were provided by Rio Tinto - Aluminium division. The company owns private weather stations (figure 1) and regularly performs snowpack measurements at 11 sites in winter. Hydrometric records are reconstructed from controlled flows from the basin outlet. The meteorological data (precipitation, maximum and minimum temperatures) are available from 1953 to the present day, as are the snowpack measurements. Snow water equivalent estimations are usually carried out manually once a month during the winter season using the average of 10 snow sampling tube measurements. Figure 2 summarizes the minimum, maximum and average daily temperatures (top panel), maximum, minimum and average monthly precipitation depths (middle panel) and historic and average streamflows (bottom panel).

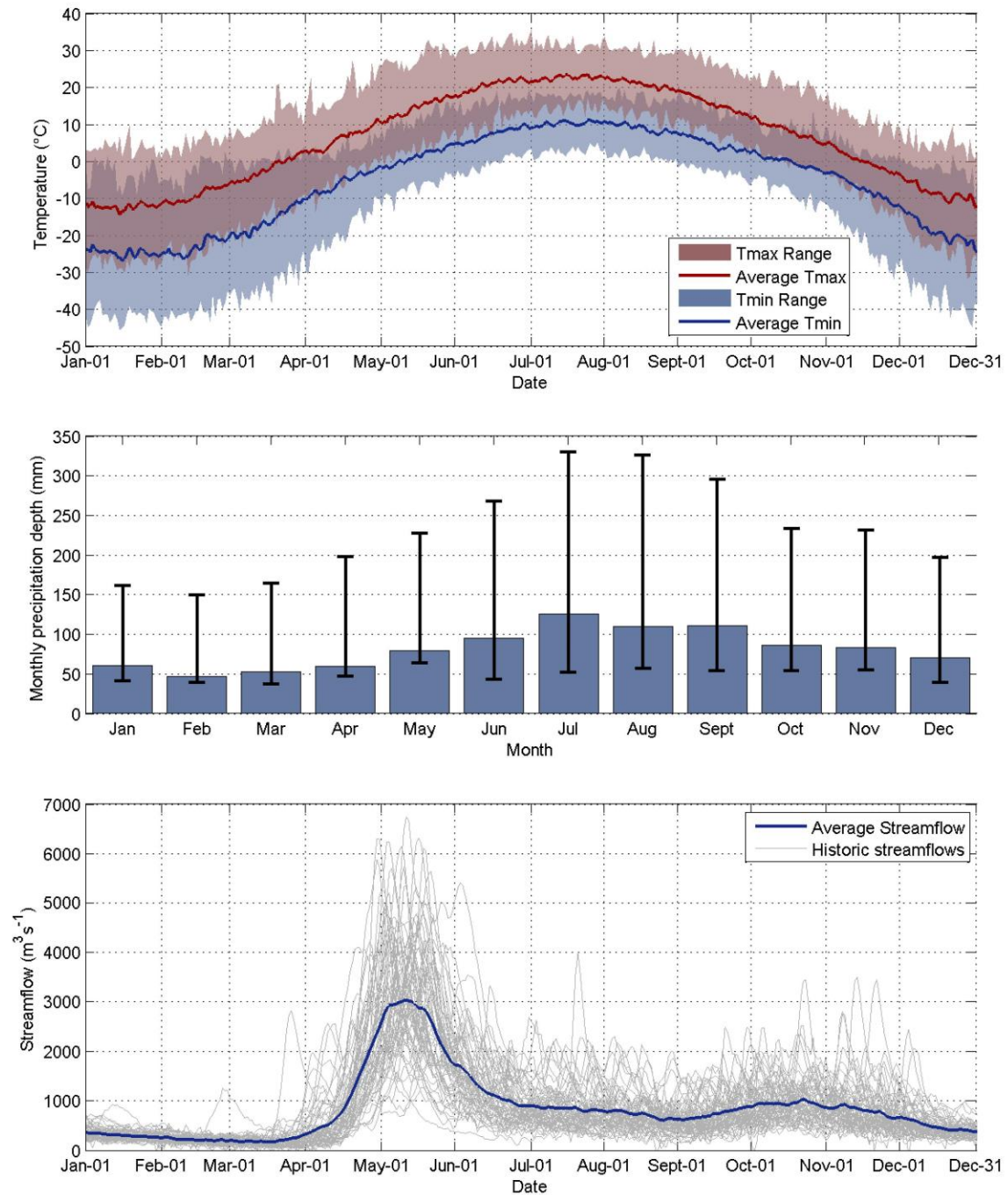


Fig. 2 Temperature (top panel), precipitation (middle panel) and streamflow (bottom panel) statistics for the 1953-2014 period on the Lac St-Jean basin

3. Methods

This section introduces the hydrological model, data assimilation and ESP approaches used in this project. The proposed HUE method is also detailed.

3.1 Hydrologic model

Rio Tinto - Aluminium division uses the CEQUEAU model operationally; therefore this study was conducted with that model (Charbonneau et al. 1977). CEQUEAU is a distributed and conceptual hydrological model which requires spatially distributed climate inputs and physical catchment characteristics. It simulates hydrologic processes on a 10km x 10km grid, including snow accumulation and melt. The model interpolates (or extrapolates as required) meteorological data to each of the grid cells which compute vertical processes independently in three vertical production stores. The flow routing module then manages the lateral displacement of water to the outlet. It can adequately take into account the variability in the spatial distribution of snowmelt during the spring, but is simple enough to execute relatively quickly. For the current study, CEQUEAU's parameters were set to their operational values thus they did not require additional calibration or validation. The method proposed in this study is independent of the model structure and its parameters, meaning that it should be applicable to any model as long as the errors are independent of the initial conditions. Therefore, the effect of using a single parameter set should have no impact on the applicability of the method and on its ability to add the appropriate amount of variability in the ensembles.

3.2 Data assimilation

Before launching ESP forecasts, one must make sure that the hydrologic model reflects the catchment's current hydrologic state. This step is usually performed by assimilating observed data into the model for a few days prior to the forecast date in order to minimize errors between the observed and simulated hydrographs. The model can then be used in simulation

mode. However, no data assimilation was performed in this study because the forecasts are only launched in winter months, during which time the model is able to simulate the low-flows relatively well. This means that the initial error is always relatively small.

3.3 Current ESP method

The method currently used in ESP forecasting for the project catchment is based on weather forecasts and resampling of the historical climate record. After the data assimilation step, the analyst samples meteorological time series from weather forecasts for the first 7 days, and then extends the forecasts with observed climate samples from past years on the same period. The analyst then feeds the forecasts to the hydrological model. For example, if a one month duration ESP is to be launched on January 1st, then weather forecasts for jan-01-jan-07 are coupled to historical climate data series for the period jan-08 to jan-31 of each year on record before being fed to the model. The generated ESP forecast is then used in the reservoir optimization algorithm. For the current study, the 7-day weather forecasts were not used. Instead, the entire climate data series were sampled from historic years. This approximation was made because of the properties of weather data in winter months on the region:

(1) Precipitation falls as snow or, more rarely, as rain that remains in the snowpack, thus simply adding to the SWE; and (2) the temperatures are overwhelmingly negative, thus 7-day forecasts have little impact on the long-term snowmelt volumes.

The problem with this method is that the ESP forecasts are often under-dispersed, leading to suboptimal decision making. Two reasons explain the under-dispersion. First, the hydrologic model, in its deterministic nature, does not reproduce the real-world variability observed when using very similar climate and/or state variables. As discussed previously, the simplification of the real-world processes leads to a reduction in output variability. Second, by using only the

years on record, there is evidently a limit on possible outcomes whereas climate sequences can be very different from the recorded observations. This project aims to correct the under-dispersion due to model structure and other sources of prediction error in such a way as to reintroduce the missing variability into the ESP forecast members.

3.4 Proposed modifications to the ESP forecast method

This paper focuses on correcting the ESP forecast under-dispersion during winter months for long-term estimation of freshet runoff volume, although the framework could potentially be applied in other periods. The efficiency of the method lies in the aggregation of the possible sources of uncertainty into a single variable. Let y^* be the actual freshet inflow volume (unknown quantity), y the measured inflow volume (measure of y^*), $\tilde{\varepsilon}_y$ the observation error on the streamflow and $M(\theta, x, I)$ the CEQUEAU model output from which the hydrological variable of interest can be calculated (here the freshet volume) with parameters θ , initial conditions x and climate inputs I . The process can be expressed as:

$$y^* = y + \varepsilon_y = M(\theta, x, I) + \varepsilon(\theta, x, I) + \tilde{\varepsilon}_y \quad (1)$$

where $\varepsilon(\theta, x, I)$ is the model error distribution which is a function of the chosen parameter set, initial conditions and climate inputs. It is usually very difficult to model the error term into individual components based on the limited observations, as was previously discussed. In this study, we neglect the streamflow measurement error $\tilde{\varepsilon}_y$ and compare the model outputs to the observed streamflow y . This leads to the simplified equation similar to that of Solomatine and Shrestha (2009) in their paper on model uncertainty estimation:

$$y = M(\theta, x, I) + \varepsilon(\theta, x, I) \quad (2)$$

In this case, we posit that the dominant hydrological variable related to the freshet runoff volume is the snow water equivalent on the catchment and that the freshet volume variability can be addressed by this state variable. This simplification allows working with a single element of the state variable vector x rather than the entire vector. Therefore the initial condition vector x was kept constant except for the SWE variable, which was used as an aggregator for the other sources of model error and state variables. In order to isolate the model's structural error, we then condition the error on the model and its parameter set, as well as for a given climate input, as shown in equation 3:

$$y_t = M(x_t|\theta, I_t) + \varepsilon(x_t|\theta, I_t) \quad (3)$$

where $M(x_t|\theta, I_t)$ is the modelled freshet volume given a parameter set θ and a climate realization I_t ; $\varepsilon(x_t|\theta, I_t)$ is the error between the measured and modelled volumes given the model's parameter and climate inputs and the index t refers to a particular year. Equation 3 differs from equation 2 in that equation 3 describes the multi-dimensional problem set, whereas in equation 3 it is possible to identify a scalar value of the model's snowpack in vector x which minimizes $\varepsilon(x_t|\theta, I_t)$, thus bringing the problem to a univariate one.

In doing so, it is possible to estimate the model error based on historical simulations. Indeed, a value of $\varepsilon(\cdot)$ can be computed for each year based on the model simulation given its parameter set and the historic climate series. This error can then be reintroduced in prediction mode.

Figure 3 illustrates the steps involved in the proposed HUE approach:

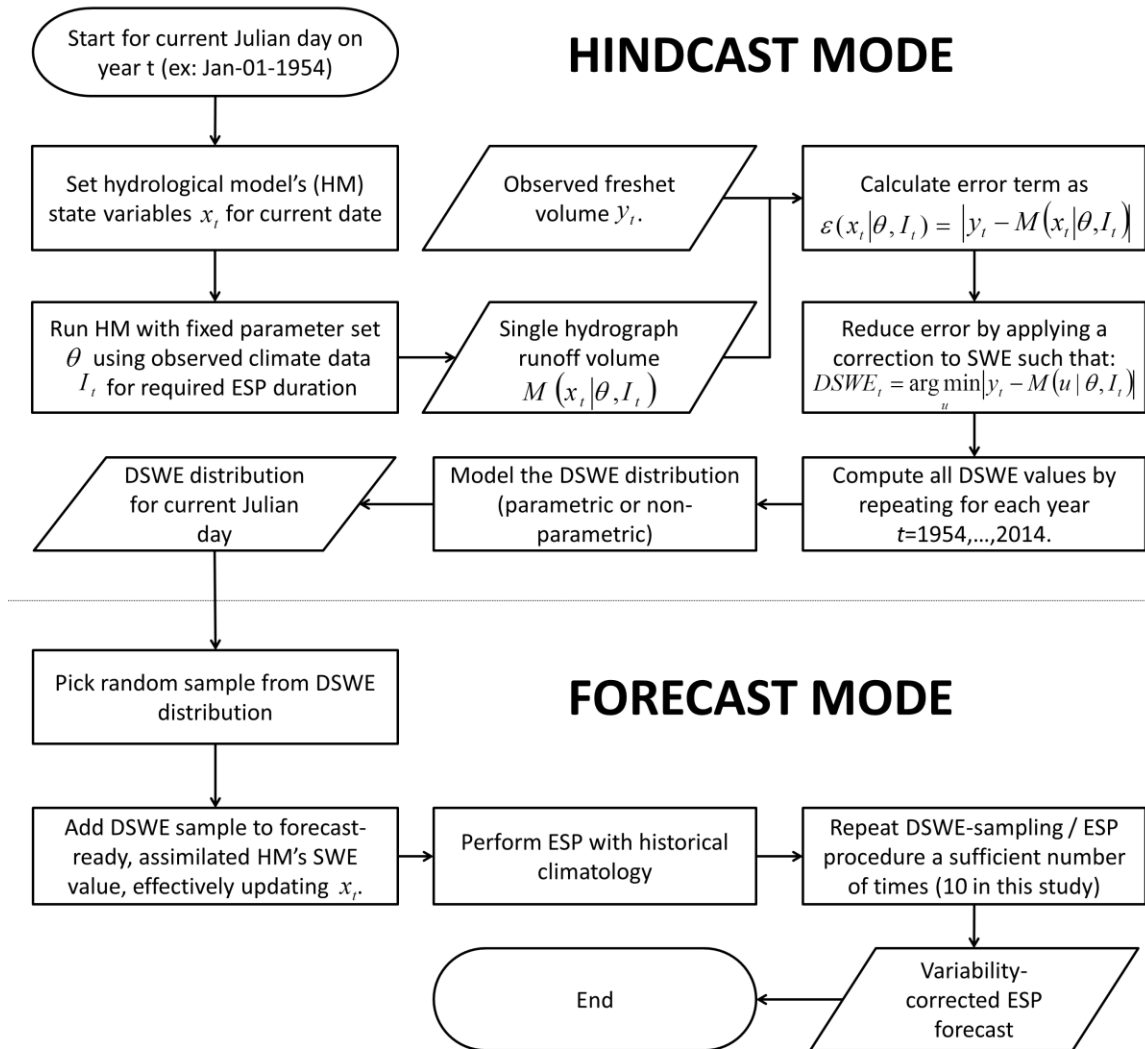


Fig. 3 Flowchart for applying the HUE method in Hindcasting and Forecasting modes

The hindcast portion in figure 3 shows how the SWE correction (DSWE value) distribution is computed. It is important to note that the process must be performed for each different forecast date to take into account the temporal aspect of the snowmelt processes. As can be seen in figure 3, there is an error between the observed and simulated runoffs even though the climate data is known (observed climate data) and the observed hydrograph is known (also

observed). The difference resides in measurement uncertainties and model structural errors.

The method allows estimating a DSWE distribution to sample from to add variability to the ESP forecasts.

The forecast mode part of the process requires sampling from the DSWE distribution and then performing a classical ESP forecast. By taking sufficient samples from the DSWE distribution and repeating the ESP forecasts, an ensemble containing many members and whose variability is corrected is generated.

This paper attempts to improve long-term volumetric forecasts, therefore performance metrics are limited to those that measure point variables. The Continuous Rank Probability Score (CRPS; Matheson and Winkler 1976), which is commonly used to validate continuous streamflow predictions, serves no use for single-variable forecasts. We therefore selected rank-type histograms as the appropriate tool to verify the ensemble forecast performance (Hamill 2001). We first order all the simulated runoff volumes from the ESP, then take the percentiles of the observed runoff volumes within the ESP volumes and then produce a histogram of these percentiles. Under-dispersion manifests itself as a U-shape histogram as the ESP forecasts are often unable to reach the higher and lower observed values due to their lack of variability, thus yielding a distribution where observed runoff values correspond to low or high percentiles (Hamill 2001). Conversely, ESPs that contain too much variability will display as a dome shaped histogram because the simulated extremes are too far compared with the observed distribution. An ideal ESP will produce a uniform histogram. More formally, the method attempts to increase the ensemble variability $Var[M(x_t|\theta, I_t)]$ so that the distribution of $\{y_1, \dots, y_n\}$ values with respect to $\{M_1^j(x_1|\theta, I_1), \dots, M_n^j(x_n|\theta, I_n), j = 1, \dots, J\}$ is uniformly distributed, where n is the number of years (thus the number of trials) on record and J is the number of ensemble members.

In this paper, three different periods are analysed, respectively beginning on Julian dates January 1st, February 1st and March 1st, each ending on June 30th which is considered as the end of the snowmelt period on the catchment. First, the traditional ESP method using historic climate sampling is run and used as a benchmark and to assess the actual ESP under-dispersion problem. Then, the proposed approach with the addition of DSWE noise to the snowpack variable is carried out and the results are compared.

4. Results and analysis

4.1 Current ESP under-dispersion

The Talagrand diagrams (percentile histograms) produced while using the traditional ESP procedure on the study catchment display signs of under-dispersion, as shown in figure 4a-4c).

Also shown in figure 4a-4c) is the Average Bin Distance to Uniformity (ABDU), which we define as:

$$ABDU = \frac{1}{N} \sum_{k=1}^N \left| s_k - \frac{M}{N} \right| \quad (4)$$

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 6.1 in this study (61 data points in 10 bins). A larger ABDU reflects a larger deviation from uniformity. It can be seen that in all cases ABDU is above 2, which means that on average, bin heights are at least 2 units away from the uniform count.

Statistical testing was used to verify if the percentiles had a distribution that significantly differed from a uniform distribution. As expected, a Kolmogorov-Smirnov test rejected the hypothesis of uniformity for all three cases at a 5% significance level (p-values = 0.007; 0.03 and 0.01, respectively).

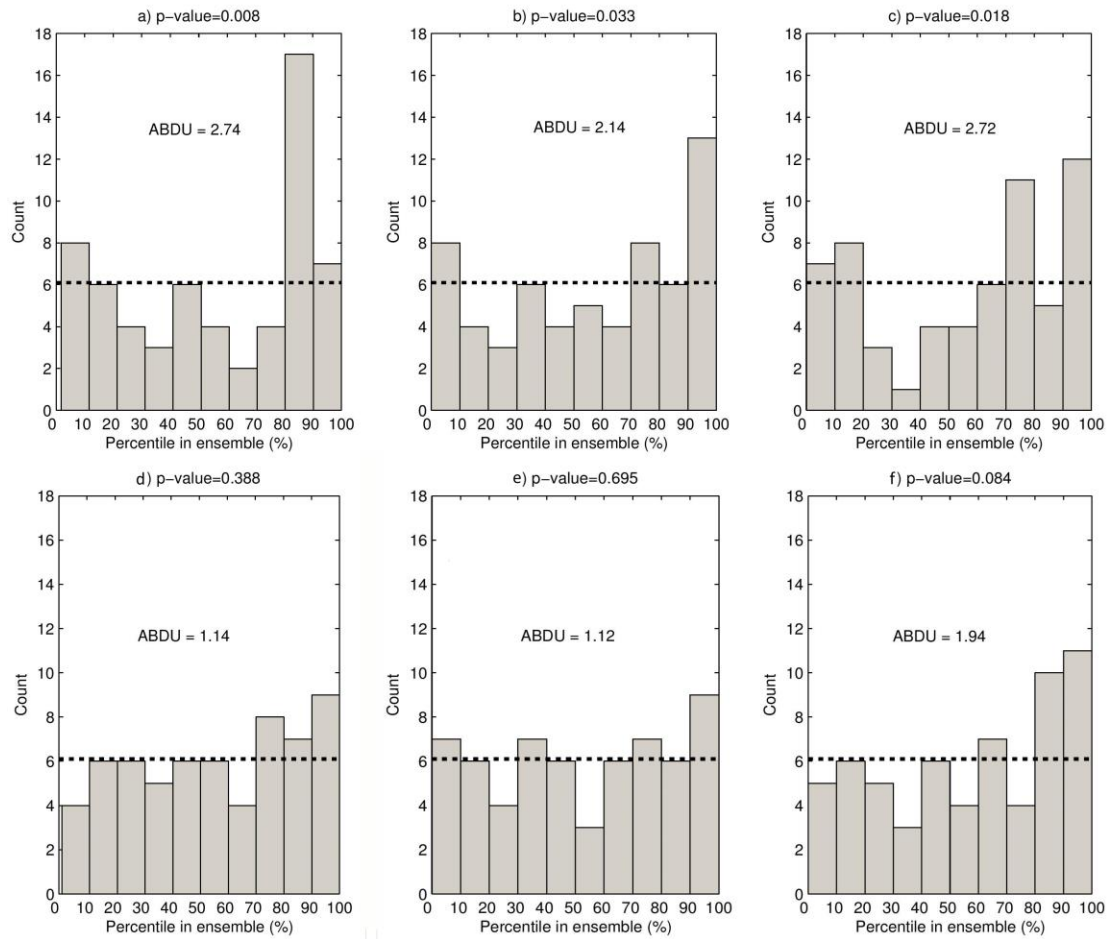


Fig. 4 Talagrand diagrams for freshet runoff volume when the forecast date is a) and d) January 1st, b) and e) February 1st and c) and f) March 1st using the traditional ESP methodology. Top of each plot: p-value of the Kolmogorov-Smirnov test for uniformity, in each case rejecting (a to c, original ESP) or failing to reject (d to f, HUE-ESP) the null hypothesis that the data are from a uniform distribution using $\alpha = 0.05$

It is clear from figure 4a-4c that the histograms display a U-shape, indicating either under-dispersion or conditional bias (Hamill 2001). We attempted to link the volume error magnitude between model simulations and other variables available at the forecast date, such as snowpack

depth, but no strong correlations were found. The evidence suggests that the problem is one of pure under-dispersion and not of conditional bias. Therefore the proposed method should allow widening the simulations to capture the more extreme runoffs.

4.2 Modeling the distribution of DSWE

The first step in the HUE approach is to estimate the change in the snowpack at the time of forecast which would lead to a perfect volumetric forecast. If the model overestimated the observations, then snow was removed from the snowpack via the hydrological model's state variables, and vice-versa. The process was repeated until the resulting error between the simulated and observed volumes was zero, i.e., find $\Delta_t : |y_t - M(x_t + \Delta_t | \theta, I_t)| = 0$. In this study, the process was automated using a simple hill-climbing algorithm (parabolic interpolation with golden section search; Brent 1973, Forsythe et al. 1977) and repeated for the same forecast date of every year (all January 1st, all February 1st, etc.). Figure 5 presents the DSWE distributions for the three forecast dates considered.

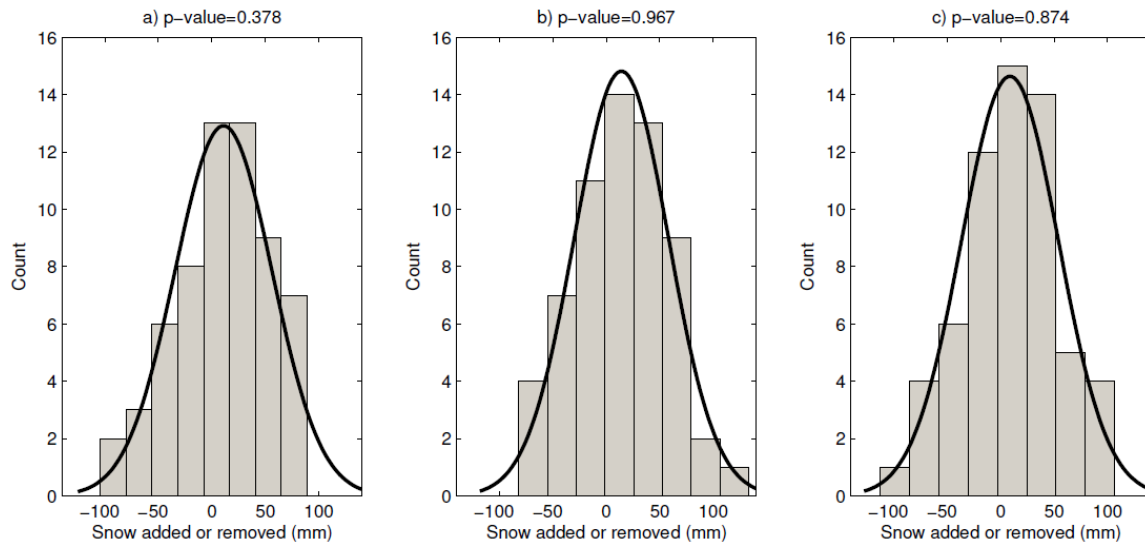


Fig. 5 DSWE distribution for the 61 years on record and for each of the three forecast dates in this study a) January 1st, b) February 1st and c) March. Overlaid is a normal distribution fitted to the underlying data. Top of each plot: p-value of the Shapiro-Wilk test for normality, in each case not rejecting the null hypothesis that the data are from a normal distribution using $\alpha = 0.05$

From figure 5, it is clear that in roughly half the cases snow must be added to the model, while for the other half snow must be removed. The distributions all display a slight positive mean with relatively large variance. This change in snow depth represents the missing model variability under the current working hypothesis. A Shapiro-Wilk test was used to determine if the DSWE distributions departed from normality. There was no evidence against normality, with respective p-values of (0.378, 0.967 and 0.874). The DSWE distributions for the three dates are respectively $\mathbf{N}(10.9, 45.4)$, $\mathbf{N}(14.2, 44.1)$ and $\mathbf{N}(8.7, 45.4)$. A linear regression between the DSWE values and the corresponding SWE state variable at forecast time, x_t , showed no correlation or trend, and also displayed normally distributed residuals. Furthermore, no correlation was found between the future snowfall amounts and temperatures in I_t and DSWE values as well. This means that the DSWE values can reasonably be generated as *iid* normal variables without regard to other variables in our study. The HUE method should nonetheless perform just as well even if the DSWE variable is not *iid*, but this remains to be validated.

4.3 Proposed HUE approach

The proposed method requires adding variability to the hydrologic model's state variables, namely SWE, during the ESP forecast. For each new streamflow trace, a DSWE value was sampled from the appropriate normal distribution as defined in the previous section. For example, ESPs launched on January 1st would require sampling from the $\mathbf{N}(10.9, 45.4)$ distribution. The sampled DSWE value is then added to the SWE state variable and the traditional ESP method is launched. This process can be repeated as many times as practically possible to insert the appropriate level of variability. A single sampling is insufficient to appropriately sample the normal distribution of DSWE values; however more samples also means more computing time. A reasonable balance was found at 10 samples for each of the 61

climate series. This means that there are a total of 610 traces instead of the initial 61 with this method. Figure 6 shows the hydrographs as simulated by the traditional ESP method (panel a) and with the HUE-ESP method (panel b) for the forecast date of January 1st 1954.

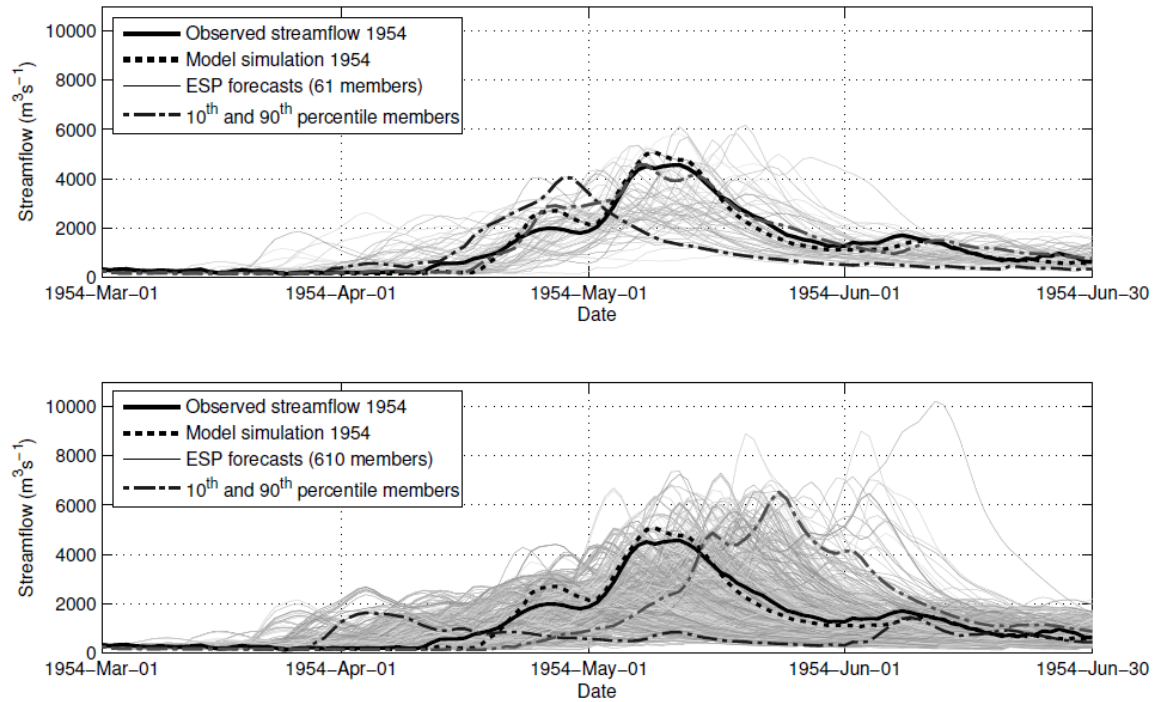


Fig. 6 ESP Forecasts launched on March 1st 1954 with the actual outcome and a posteriori modelled outcome using the traditional ESP (top panel) and with DSWE resampling (bottom panel). Also shown are the members representing the 10th and 90th percentiles of predicted freshet runoff.

In each case, the actual realization is also presented (note that this realization is unknown at forecast time). The model simulation given the actual observed climate data is also shown, as are the 10th and 90th percentile hydrographs. The corrected ESP method displays more variability than the original ESP approach, which means that in cases where the original ESP narrowly misses the observed volume, the corrected ESP method is more likely to capture it due to the added variability. It is clear that the peak flow timing is not substantially modified, which is normal since the melt is dependent on the temperature, which in turn is obtained with the same 61 historical climate series with the two ESP approaches.

Using this increased-variability approach, Talagrand diagrams were produced for the same forecast dates as with the original ESP method, which were January 1st, February 1st and March 1st. Figure 4d-4f) shows the results for the proposed method.

As compared to the traditional ESP method, it can be seen that the ABDU values were reduced considerably, with January and February forecasts seeing reductions on the order of 50%. For both these cases, there is also no evidence against the uniformity hypothesis using the Kolmogorov-Smirnov test, with p-values of 0.388 and 0.695. For the March forecast, the ABDU value was reduced from 2.72 to 1.94, a 28% reduction. However, the Kolmogorov-Smirnov test shows weak evidence against uniformity, with a p-value of 0.084. This can be explained by the fact that for a few of the years on record, snowmelt starts in early to mid-March, and state variables other than snow begin to have more importance in the model. By exclusively modifying the snowpack, the method loses some of its potential during this period. Nonetheless, the results have improved markedly compared with the traditional ESP approach.

5. Discussion and analysis

The method presented in this paper showed promising results on the study catchment. However it does have some drawbacks which could be problematic for certain applications. In this section, we discuss its strengths and weaknesses.

5.1 Review of the method's performance

When comparing the proposed method's performance to that of traditional ESP forecasting, our method is clearly able to generate acceptable results for use in reservoir optimization applications, whereas the traditional ESP often generates under-dispersed ensembles. However it is important to remember that the objective in this paper was to improve seasonal volumetric

forecasts, therefore peak flood, timing and other time-dependent variables were neglected. Furthermore, by the very nature of the method used to add variability (modifying the model's SWE initial conditions), it is not possible to control the timing of the inflows. However, by applying the method with other state variables such as the snowpack temperature index, it could be possible to take the timing into account, but this remains to be validated.

Other studies focusing on volumetric forecasts were also able to improve under-dispersion of the ensemble. DeChant and Moradkhani (2014) show that adding uncertainty to the initial conditions improves the ensemble distribution tails, whereas model structure uncertainty improves the medial portion of the ensemble. However their method relies on using multiple models and data assimilation techniques to estimate the distribution of initial conditions instead of directly modeling the error distribution. The Bayesian model averaging scheme they used is also complex and time-consuming for larger distributed models, which can be a limiting factor in an operational setting.

Najafi and Moradkhani (2015) also arrive at the same conclusions using multi-model combination techniques. Once again, the use of multiple models is not necessary in the proposed approach because the DSWE already contains the model structural error. Instead of isolating each source of uncertainty, we introduce all uncertainty into the model through the initial states. Furthermore, using multiple models is not appropriate for generating ESPs for feeding a reservoir optimization algorithm because doing so would make it difficult to isolate and keep track of the transition variables between hydrologic values from the model and the ESPs. Given these constraints, the proposed HUE method seems the most efficient.

5.2 DSWE analysis

The histograms in figure 5 show that there is a slightly positive mean in the DSWE distributions. A one-sample t-test determined that the means were not statistically different from zero at the 5% level for the months of January and March, with p-values of 0.06 and 0.13 respectively, whereas the month of February returned a p-value of 0.01. However the small p-value for January indicates that the distribution mean is bordering the significance threshold. In any case, this implies that snow must be added to the model more often than the opposite. A few reasons may explain this behaviour. First, the model uses precipitation measured from a series of gauges spread over the catchment. It is well known that gauges can underestimate precipitation values due to wind (WMO 2008). The effect is more pronounced during winter, as snow may accumulate and block the top of the gauge, which prevents further snow from entering and being accounted for. Therefore it is expected that snow might need to be added to the model to compensate for this type of error during volumetric forecasts of freshet runoff.

Second, the way DSWE is estimated requires being able to completely remove the difference between simulated and observed runoffs. However, in some cases (especially early in the season when the snowpack is small), it is impossible to remove enough snow from the model to reduce future discharge to the observed level. In such cases, DSWE was set to the value that reduced the snowpack level to zero. This leads to a positively biased DSWE for that year. The DSWE was not discarded from the distribution because the fact that it is negative still brings important information, albeit incomplete.

While the DSWE distribution's mean is not quite zero, its variance is much more important. This can be interpreted as the processes being mostly correct but failing to encompass the more

extreme scenarios. The Talagrand diagrams in figure 4 show that this lack of variability was addressed and corrected adequately with the HUE method.

Furthermore, although not shown in this paper, the method has been successfully implemented on four other sub-basins in the 75,000 km² Lac-St-Jean basin as well as on another basin (Kemano reservoir basin) in British-Columbia, Canada, for medium-term freshet volume forecasting.

5.3 Limitations

There are a few limitations with the HUE method. As was discussed previously, it is designed to improve uncertainty estimation for single hydrological variables; therefore it cannot be used for time-dependent variable estimation, such as runoff time series. This must be kept in mind in order to adequately assess the method's limitations.

First, to be effective, the study site is required to have relatively long climate and streamflow records. Longer records allow for a better estimation of the DSWE distributions. If only 5 or 10 years of data are available, then the DSWE distributions will not be robust. Following this, data assimilation must also be of good quality to ensure proper initial conditions are set in the hydrological model. Although true for traditional ESPs also, poor initial conditions in the model could bias the results and lead to inappropriate DSWE values.

Second, in this study, the seasonal forecast was performed during the winter which allowed using the snowpack as an integrator of all sources of uncertainty. However, when there is no snowpack (warmer catchments, summer season), other state variables must be used. In these cases, volumetric streamflow forecasting is more of a challenge because of the increased stochasticity of the hydrological processes. For example, soil water levels may change drastically

based on weather patterns. Furthermore, uncertainty in the future meteorological conditions is larger during warm periods due to the added influence of temperature. During winter, the cold temperatures limit the amount of variability that can be induced from the climate data. These findings are in line with Roulin and Vannitsem (2015) who showed that errors in climate forcings are more important than initial conditions except for the winter season.

5.4 Distribution fitting

During the DSWE distribution fitting, it was found that the normal distribution adequately represented the DSWE values. However, this might not always be the case. In such situations, non-parametric methods, such as kernel density estimation, could be used instead. A test was performed using normal kernels and a bandwidth value which minimizes the Asymptotic Mean Integrated Square Error (AMISE; Jones et al. 1996). The convoluted distribution was nearly identical to the normal distribution and produced similar results (not shown). Preliminary tests during summer rainfall-dominated periods showed promising results using a non-parametric approach. Therefore it could be possible to adapt the method to non-normal DSWE distributions; however the efficacy of this approach remains to be validated. Furthermore, in other conditions or with other models, $\Delta(x_t/\theta, I_t)$ (or DSWE) might be dependent on initial conditions in vector x_t or on the climate inputs I_t , which would require more complex modelling than the one in this study.

6. Conclusions

In this paper, we propose a novel method (HUE) to correct ensemble streamflow prediction under-dispersion for use in reservoir optimization applications on snowmelt-dominated catchments. The approach relies on integrating all sources of uncertainty into a single initial condition variable - in this case the snow water equivalent in the hydrological model. This error

was estimated for each year and its distribution was modelled. Then, when performing ESP forecasts, variability was added to the initial conditions by sampling from this distribution..

One of the main advantages of adding variability to the initial conditions is that every trace can be linked to the hydrological model's state variables and climate data, which allows computing appropriate transition variables for stochastic dynamic programming optimization of reservoir water levels. There are many published papers that address under-dispersion in volumetric streamflow forecasting, however they mostly use post-processing methods and multi-model techniques. These make it impossible to identify transition variables during the optimization process. However, for applications that do not have this requirement, alternative methods such as ensemble post-processing might return better results.

It is important to restate that the overall uncertainty in the hydrological variable prediction comes from many sources, namely the climate inputs, the model structure and the state variables. In this project we bundled the initial conditions and model structural error into a single term because it was the most efficient approximation for the winter period. The error was conditioned on model response to observed climate series and it was found that the errors were uncorrelated to the runoff volumes. However the climate inputs also play a role in the overall uncertainty in hydrological prediction. The other facets of the uncertainty will be subject to further research to improve the overall uncertainty estimation.

On the Canadian catchment in this study, the results showed that it was possible to correct the ensemble under-dispersion for lead-times of 4 to 6 months as compared to traditional ESP forecasting. Kolmogorov-Smirnov tests confirmed that there was significant improvement in the ESP forecasts after adding the uncertainty to the initial conditions. The under-dispersion was eliminated in all cases, with the Talagrand diagrams showing more uniform distributions. While

the method was only tested on a handful of basin during winter, the concept should be applicable to other sites and conditions. Further research should focus on adapting the approach to other types of initial conditions and combinations thereof.

7. References

- Arsenault R, Malo J-S, Brissette F, Minville M, Leconte R (2013) Structural and non-structural climate change adaptation strategies for the Péribonka water resource system. *Water Resour Manag* 27(7): 2075–2087.
- Boucher M-A, Anctil F, Perreault L, Tremblay D (2011) A comparison between ensemble and deterministic hydrological forecasts in an operational context. *Adv Geosci* 29:85-94. doi:10.5194/adgeo-29-85-2011.
- Boucher M-A, Perreault L, Anctil F, Favre A-C (2015) Exploratory analysis of statistical post-processing methods for hydrological ensemble forecasts. *Hydrol Process* 29(6) : 1141-1155.
- Brent RP (1973) *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Bröcker J, Smith LA (2008) From ensemble forecasts to predictive distribution functions. *Tellus A* 60: 663–678.
- Charbonneau R, Fortin J-P, Morin G (1977) Le modèle CEQUEAU: Description et exemples d'utilisation dans le cadre de problèmes reliés à l'aménagement [Description and examples of its use in problems related to water resource management]. *Hydrol Sci Bull* 22(1): 193–202.
- Chen J, Brissette FP (2015) Combining stochastic weather generation and ensemble weather forecasts for short-term streamflow prediction. *Water Resour Manage* 29(9) : 3329-3342.
- Chen J, Brissette FP, Li Z (2014) Postprocessing of Ensemble Weather Forecasts Using a Stochastic Weather Generator. *Mon Wea Rev* 142 : 1106–1124. doi:10.1175/MWR-D-13-00180.1
- Clark MP, Hay LE (2004) Use of medium-range numerical weather prediction model output to produce forecasts of streamflow. *J Hydrometeorol* 5: 15-32.
- Côté P, Leconte R (2015) Comparison of Stochastic Optimization Algorithms for Hydropower Reservoir Operation with Ensemble Streamflow Prediction. *J Water Resour Plann Manage*. doi:10.1061/(ASCE)WR.1943-5452.0000575, 04015046.
- Côté P, Haguma D, Leconte R, Krau S (2011) Stochastic Optimization of Hydro-Quebec hydropower installations: a statistical comparison between SDP and SSDP methods. *Can J Civil Eng* 38(12): 1427-1434.

Day GN (1985) Extended streamflow forecasting using NWSRFS. *J Water Resour Plann Manage* 111:157–170.

DeChant C, Moradkhani H (2011) Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation. *Hydrol Earth Syst Sc* 15: 3399–3410. doi:10.5194/hess-15-3399.

DeChant C, Moradkhani H (2014) Toward a reliable prediction of seasonal forecast uncertainty: Addressing model and initial condition uncertainty with ensemble data assimilation and Sequential Bayesian Combination. *J Hydrol* 519(D) : 2967–2977.

Desreumaux Q, Côté P, Leconte R (2014) Role of hydrologic information in stochastic dynamic programming: a case study of the Kemano hydropower system in British Columbia. *Can J Civil Eng* 41(9):839–844.

Faber BA, Stedinger JR (2001) Reservoir optimization using sampling SDP with ensemble streamflow prediction (ESP) forecasts. *J Hydrol* 249(1–4) : 113–133. doi:10.1016/S0022-1694(01)00419-X.

Forsythe GE, Malcolm MA, Moler CB (1977) *Computer Methods for Mathematical Computations*, Prentice-Hall, Englewood Cliffs, New Jersey.

Fortin V, Favre A-C, Saïd M (2006) Probabilistic forecasting from ensemble prediction systems: improving upon the best-member method by using a different weight and dressing kernel for each member. *Q J Roy Meteor Soc* 132(617): 1349–1369.

Franz KJ, Hogue TS, Sorooshian S (2008) Snow model verification using ensemble prediction and operational benchmarks. *J Hydrometeorol* 9:1402–1415.

Gneiting T, Raftery A-E, Westveld A-H, Goldman T (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon Wea Rev* 133(5): 1098–1118.

Hamill TM (2001) Interpretation of Rank Histograms for Verifying Ensemble Forecasts. *Mon Wea Rev* 129: 550–560. doi:10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.

Hashino T, Bradley AA, Schwartz SS (2007) Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrol Earth Syst Sci* 11 : 939–950. doi:10.5194/hess-11-939-2007.

Jones MC, Marron JS, Sheather SJ (1996) A brief survey of bandwidth selection for density estimation. *J Am Stat Assoc* 91(433): 401–407.

Liu Y, Gupta HV (2007) Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resour Res* 43: W07401. doi:10.1029/2006WR005756.

Liu Z, Guo Y, Wang L, Wang Q (2015) Streamflow forecast errors and their impacts on forecast-based reservoir flood control. *Water Resour Manage* 29(12): 4557–4572.

Matheson JE, Winkler RL (1976) Scoring rules for continuous probability distributions. *Manage Sci* 22(10) : 1087–1096.

Najafi MR, Moradkhani H, Piechota TC (2012) Ensemble Streamflow Prediction: Climate signal weighting methods vs. Climate Forecast System Reanalysis. *J Hydrol* 442–443: 105–116. doi:10.1016/j.jhydrol.2012.04.003.

Najafi MR, Moradkhani H (2015) Ensemble Combination of Seasonal Streamflow Forecasts. *J Hydrol Eng*. 10.1061/(ASCE)HE.1943-5584.0001250, 04015043.

Pagano TC, Shrestha D, Wang Q, Robertson D, Hapuarachchi P (2013) Ensemble dressing for hydrological applications. *Hydrol Process* 27(1): 106–116.

Poulin A, Brissette F, Leconte R, Arsenault R, Malo JS. (2011) Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. *J Hydrol* 409(3–4): 626–636.

Raftery A, Gneiting T, Balabdaoui F, Polakowski M (2005) Using Bayesian model averaging to calibrate forecast ensembles. *Mon Wea Rev* 133(5) : 1155–1174.

Roulston MS, Smith L-A (2003) Combining dynamical and statistical ensembles. *Tellus A* 55 :16–30.

Roulin E, Vannitsem S (2015) Post-processing of medium-range probabilistic hydrological forecasting: impact of forcing, initial conditions and model errors. *Hydrol Process* 29: 1434–1449. doi: 10.1002/hyp.10259.

Sloughter JM, Raftery AE, Gneiting T (2006) Probabilistic quantitative precipitation forecasting using Bayesian model averaging. Tech Rep 496, Dep. of Stat., Univ. Of Wash., Seattle.

Solomatine DP, Shrestha DL (2009) A novel method to estimate model uncertainty using machine learning techniques. *Water Resour Res* 45: W00B11. doi:10.1029/2008WR006839.

Sproles EA, Kerr T, Nelson CO, Aspe DL (2016) Developing a Snowmelt Forecast Model in the Absence of Field Data. *Water Resour Manage* 30(7): 2581–2590.

Troin M, Arsenault R, Brissette F (2015) Performance and Uncertainty Evaluation of Snow Models on Snowmelt Flow Simulations over a Nordic Catchment (Mistassibi, Canada). *Hydrology* 2(4): 289–317. doi:10.3390/hydrology2040289.

Van Schaeybroeck B, Vannitsem S (2011) Post-processing through linear regression. *Nonlinear Proc Geoph* 18(2) : 5147–5160.

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. *Adv Geosci* 29 : 33–42.

Vrugt JA, Gupta HV, Nualláin BO, Bouten W (2006) Real-Time Data Assimilation for Operational Ensemble Streamflow Forecasting. *J Hydrometeor* 7: 548–565. doi:10.1175/JHM504.1.

Wang X, Bishop C (2005) Improvement of ensemble reliability with a new dressing kernel. Q J Roy Meteor Soc 131(607): 965-986.

World Meteorological Organization (WMO) (2008) Guide to meteorological instruments and methods of observation. WMO-8 8:1-681.

Zalachori I, Ramos M-H, Garçon R, Mathevet T, Gailhard J (2012) Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. Adv Sci Res. 8: 135-141. doi:10.5194/asr-8-135-2012.

Zeinivand H, De Smedt F (2009) Hydrological modeling of snow accumulation and melting on river basin scale. Water Resources Manage 23(11): 2271-2287.