Authors' response for the manuscript "Combined assimilation of streamflow and snow water equivalent for mid-term ensemble streamflow forecasts in snow-dominated regions" by J. M. Bergeron, M. Trudel and R. Leconte

We would like to thank the referees for providing constructive comments for the manuscript. Their suggestions paved the way for many changes in the revised version of the manuscript. The main changes introduced are the following:

1) A rearrangement of information on the generation of synthetic observations and meteorological input, as well as their ensemble versions, into individual sections. Additional clarifications were also added.
2) A section specifically on the generation of ensemble streamflow predictions.
3) Additional comparisons of results with other studies that share some similarities in section 4.1.1, 4.2.1 and 4.2.2.
4) A new figure comparing the error distributions used with their analogous normal distributions obtained using similar parameters. Justifications for the use of those distributions were added to the sections where they are introduced (sections 4.1.1 and 4.1.2)

Many other small changes were made as suggested by the referees. These changes can be seen in the marked-up version of the manuscript with changes tracked below, following the point-by-point response to each of the referees.

Italicized text: Referee's comment

AR: Authors' response

Comments to Referee #1

In this study, the authors explore the assimilation of discharge, SWE and SCA in a hydrologic model for the potential to improve streamflow forecasting in a mountainous basin in western Canada. Synthetic data sets are developed and used. The authors first determine which state variables are adequately predicted by the three data types that are candidate for assimilation. SCA was found to not be a good predictor. Then, the impact of assimilating SWE and discharge on hydrologic forecasts was tested.

Overall, this is an interesting study with good results. Forecasts were improved with SWE and Q assimilation both when assimilated individually and simultaneously. It is demonstrated that the data were useful for adjusting several model states (VOL, GWL, and SWE) in the CEQUEAU model. The methods in this paper show promise for applications in forecasting provided the results remain consistent for non-synthetic studies.

General comments:

1) There needs to be more detail provided in the methods section. As is, it appears as though the methods are valid, but I could not replicate this study with the information provided. I find myself having to assume I know what the authors did during some steps. Therefore, specific comments about where to add necessary detail are provided below.

AR: Information has been rearranged into new sections and additional details were added as recommended in the specific comments section. Section 3.1, dealing with the overall approach used during the synthetic experiment, now has three subsections explaining specifically how synthetic observations (3.1.1), synthetic meteorological input (3.1.2) and ESPs (3.1.3) were generated. Sections 3.2.2 and 3.2.3 now deal exclusively with meteorological ensemble and observation ensemble generation, respectively. Some information may appear to be redundant at first, but this is because similar approaches were used to generate perturbations required in the creation of single-valued synthetic observations/meteorological input and their ensemble version.

2) In the results section, the authors should make a stronger effort to link their findings to other studies. There are several papers referenced that explore assimilation of SWE and/or discharge in snow-dominated areas. There are also likely studies that have examined this type of data in other modeling and forecasting contexts. While there are a few comments about results from other studies, the authors should try to add more to the discussion.

AR: Comparisons with other studies have been added in multiple sections, including the state vector configuration when assimilating streamflow (4.1.1) and the impact of streamflow (4.2.1) and SWE (4.2.2) data assimilation on streamflow forecasts. Additional references will also be introduced for comparison purposes.

Specific comments:

Page 2, line 30. The last sentence might be better as a statement rather than a question. It seems out of character with the rest of the introduction.

AR: The sentence has been reformulated as a statement:

"The importance of state vector configuration when using multivariate DA for hydrological modeling has yet to be investigated."

Page 3, line 9: "such that the difference in elevation reaches about 1700m" is oddly stated. The difference in elevation between what? If 1700m is the total relief of the mountain, simply state it that way.

AR: The difference in elevation between the highest and lowest point in the watershed reaches about 1700m. This has been clarified in the revised manuscript.

Page 4, line 8: US Army Corps of Engineers should be capitalized. Also, is the inclusion of "1956" intended to be a citation? There is nothing listed in the references regarding this.

AR: The missing reference has been added and capitalized.

Page 4, line 17: SI is not in equation 1. How is it relevant to this discussion?

AR: SI is in the denominator of equation 1. It is one of three parameters used to convert snow water equivalent into snow cover area. It is included, along with other parameters, to be transparent in our approach and help readers understand the method used to produce the results shown in the manuscript.

Page 4, Line 18-19: As written it is implied that Hall et al. (2002) calibrated the three parameters in Equation 1. I do not think that is the case. Additional explanation of this calibration is needed. Who conducted the calibration? Was it conducted for this region? If not, is it considered to be universally applicable?

AR: The reference was initially meant to be a reference for the MODIS data, but this was not clear from the way the sentence was structured.

All the model parameters were not calibrated in the same way. CEQUEAU has numerous parameters that have been manually calibrated by engineers working for Rio Tinto, our industrial partner. These include snow, soil, evapotranspiration and transfer parameters. However, since snow cover area is not explicitly computed by CEQUEAU, a depletion curve had to be appended. The details pertaining to the depletion is not found in CEQUEAU's user manual or in any other study using or detailing how CEQUEAU works. To be transparent in our approach, we explain how the depletion curve is computed, requiring three parameters. The calibration of those parameters was conducted using the SCE-UA method (Duan et al, 1992) to minimize the root mean square difference between simulated snow cover area and MODIS data for each whole square within the Nechako watershed. The parameters are therefore not likely to be "universally applicable", but the approach may be.

The sentence has been reworked in the revised version of the manuscript to include more information about the calibration process.

Page 4, line 20: Tampered not tempered.

AR: This has been corrected in the revised manuscript for all occurrences of the term.

Page 6, line 6-8: this statement is becoming repetitive. It was mentioned several times in this section that it has been shown useful in hydrologic studies. I recommend removing earlier statements like this, or combine them into one or two sentences.

AR: The sentence has been modified to better transition into the following section.

Page 6, lines 14-24: It isn't clear what variables are referred to when using the term "observations". This may be stated earlier, but it would help the reader if they were explicitly stated here.

AR: The term "observations" used in the manuscript refers to observations to be assimilated using the EnKF, namely streamflow, SWE and SCA. The expression "meteorological input" or "weather input" is used to refer to precipitation and mean air temperature. There is one occasion in the manuscript where the expression "meteorological observation" is used and has been corrected to avoid confusion. Additional clarifications have also been added to explicitly state what are "observations" and "meteorological input".

In general, this section lacks detail. In what way and by how much were the data perturbed? How do you get synthetic observations by perturbing "true states"? More specific terminology and combining or pulling in information from Section 3.2 would be helpful in understanding the procedure of creating synthetic data.

AR: Initially, sections 3.2.2 and 3.2.3 described how both the individually perturbed and ensemble version of meteorological input and observations were generated since they both used a similar approach using the same perturbation factors. The information was added in the "Hyper-parameter tuning" section since it partly dealt with specifying errors, which is required when generating ensembles.

However, this could lead to confusion between the two sets and appeared to create a void in the description of the synthetic experiment. To avoid confusion, the information from sections 3.2.2 and 3.2.3 strictly dealing with initial perturbations of observations and meteorological input have been used to create additional subsections (3.1.1 and 3.1.2) following the overview of the synthetic experiment.

The methods section includes very little description of how ESP forecasts are generated. A more thorough explanation should be provided for readers unfamiliar with the process. Please clarify whether only 20 years of meteorological data (1990-2000) were used to generate the ESP ensembles. Also, was only the mean value of the state variable predicted by the EnKF used to generate each ensemble in the ESP forecast, or were multiple state values from the state variable ensemble used?

AR: ESPs were generated everyday over the entire study period (10 years) using a forecast horizon spanning 50 days. The multiple state values resulting from the EnKF were used to generate ESPs and is the only factor differentiating ensemble members during the forecast phase. The same true meteorological input and model parameters were during the forecast phase. No meteorological ensemble was used to generate ESPs. Although this generates forecasts which are not "realistic" since they get better over the forecast horizon, this is a way to evaluate the potential impact of data assimilation on streamflow forecasts, which is one of the two goals of the study. Using a meteorological ensemble forecast would simply add unnecessary noise. An additional section (3.1.3) has been added on ESP generation.

Page 10, line 24 onward: What is the timestep of the data evaluated? Hourly, daily, etc? I cannot find where this is explicitly stated, but it is important to understanding the results of the study. If the streamflow is evaluated at a daily timestep, it makes sense that the SWE is not beneficial for predicting streamflow; however, results might be quite different if output is evaluated at a monthly or seasonal timestep. In addition, I could not find the interval between assimilation, is it done at each model timestep, daily, weekly?

AR: The time step is daily. This has been clarified in the revised manuscript by mentioning a daily time step for the model and daily availability for the observations.

Page 14, lines 1-6 and Figure 8: It is not quite clear what VOL is in the model. As presented on page 4, it appears to be water that is being routed to the outlet (i.e. runoff). If that is the case, the quick decline in adjusting the VOL state on the CRPSS makes sense not because of a linear relationship with discharge, but because of the likely short residence time of the water represented by VOL within the watershed. The authors discuss the residence time issue in the next paragraph with respect to GWL and SML, I would like to see similar insight regarding the VOL as the linear relationship explanation is not obvious.

AR: The initial impact strength is due to the close link between streamflow and VOL. Since the correlation between the streamflow observation and VOL is relatively high, VOL globally experiences a positive update (as seen in Fig. 4). In return, since simulated streamflow depends strongly on VOL, simulated streamflow initially experiences an important positive impact. However, the duration of this impact is indeed caused by the water residence time. A short-lived impact on CRPSS as a function of forecast horizon would be caused by a relatively short residence time. This was not obvious from the way it was stated in the initial manuscript and has been clarified. Additional clarifications have also been added to section 2.2 concerning the role of VOL and streamflow.

Page 16, lines 14-16 and Figure 10: It would be helpful to add a sentence putting results from Figure 10 in context of results from assimilating only Q or only SWE.

AR: A few sentences have been added to address the comparison between the combined and individual data assimilation scenarios:

"CRPSS values for combined assimilation of both streamflow and SWE observations were superior to CRPSS values for individual assimilation of streamflow or SWE over all forecast horizons, with the exception of forecast horizons higher than 45 days, where CRPSS values for SWE DA are slightly higher. This reveals that while the updated VOL and GWL by streamflow data assimilation may be very beneficial for short-term forecasts, they do not further improve the mid-term forecasts when combined with SWE data assimilation in comparison with the scenario where only SWE data is assimilated."

Page 17, lines 28-29: The SCA was not tested on the forecasts due to the lack of improvements in state variables (page 13, lines 19-20). The authors should not make any conclusions regarding the impact on forecast skill from this study, and restate this with respect to state variable improvements.

AR: This has been corrected in the revised manuscript.

Page 18, line 4: The statements made in this paragraph are not hypotheses, they are assumptions and limitations

AR: This has been corrected in the revised manuscript.

References :

Duan, Q. Y., Sorooshian, S. and Gupta, V.: Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models, Water Resour. Res., 28(4), 1015–1031, doi:10.1029/91wr02985, 1992.

Comments to Referee #2

The manuscript titled Combined assimilation of stream flow and snow water equivalent for mid-term ensemble stream flow forecasts in snow-dominated regions is an attempt to apply the Ensemble Kalman Filter to the application of stream flow prediction in regions where the majority of the water involved comes from snow.

The manuscript undertakes a set of studies, the first is to ascertain which of the 7 possible state variables they consider are sensitive to the assimilation of the 3 observation types they consider. They indicate that snow cover appears to not be an important factor in the forecasts that they seek.

The manuscript is well written and upon a second reading easy to follow. However, i do have a couple of points that need to be addressed before i can sign off on publication

Major Comments:

1) My first concern relates to the generation of the ensemble perturbations and the perturbations to the observations. You indicate that you use Gamma, lognormal of beta distributions yet the EnKF is highly reliant on these perturbations, and hence the errors being Gaussian distributed. My query, and i am requesting graphs of these, is to see the distributions plots for the distributions that you mentions with the parameters in the manuscript. My hunch is that these will look quite close to a Gaussian distribution of some form and as such is why you obtain the results, which are great results, but it could be misleading to have these distributions when really they are close to a Gaussian.

AR: Many studies skip the controlled experiment and go straight to the assimilation of real data, but we feel this is not the ideal approach. As a first step to test DA potential on streamflow predictions, a near-ideal framework should be constructed in order to reduce the number "outside variables" that can influence results and mislead the analysis. Since the EnKF is used, all perturbations should have a normal distribution in order to obtain optimal results. However, observations like the ones used in our study have physical limits that cannot be breached (eg: SWE must be >= 0). Alternatives approaches must be used. In our study, we decided to use different distributions that 1) resemble normal ones when the mean is away from the limits, 2) prevent any violation of the physical limits of the variable and 3) are unbiased. The distributions introduced in the manuscript fit that description.

Examples of beta distributions obtained using the parameters presented in the manuscript, along with their analogous normal distributions obtained using the same mean and variance, are shown below (Fig. R1a of this response, has been added as Fig. 4a in the revised manuscript). The variance is defined as a function of the mean in such a way that it is largest at 0.5 and smallest at the extremes in an attempt to reflect MODIS SCA retrieval's greater uncertainty during the transition periods when patchy snow is prominent. The beta distribution also prevents perturbations from violating the physical limits of the variable, which the normal distribution cannot guarantee. A visual glance at the graphs shows that the beta distributions resemble normal distributions most near 0.5 and deviate more as the mean gets closer to the extremes.

In a similar fashion, examples of gamma and lognormal distributions using the parameters (variance values) described in the manuscript are show in Fig. R1b of this response (has been added as Fig. 4b in the revised manuscript). A mean of 1 has been used for all examples for an easier comparison between examples, but the distributions themselves are visually independent of the mean since the variance is defined as directly proportional to the mean. The distributions resemble most a normal one for small variances relative to the mean, but important deviations can be noted as the relative variance increases due to the lower limit (zero) imposed on gamma and lognormal distributions. The variances shown are the ones used to perturb SWE and streamflow observations, as well as meteorological input.

While those distributions may or may not reflect entirely the real error of the observations and meteorological input, therefore yielding "optimistic" results, we feel they are a good compromise between the "normality" required by the EnKF and physical limits of the variables. As future works, one could investigate other distributions, introduce a bias, etc. and analyse their impact, but this falls outside the scope of the current manuscript.

The assumptions and limitations of the method are stated in the conclusion. In order to make it clearer that the experiment is near ideal, an additional sentences has been added in the revised manuscript.
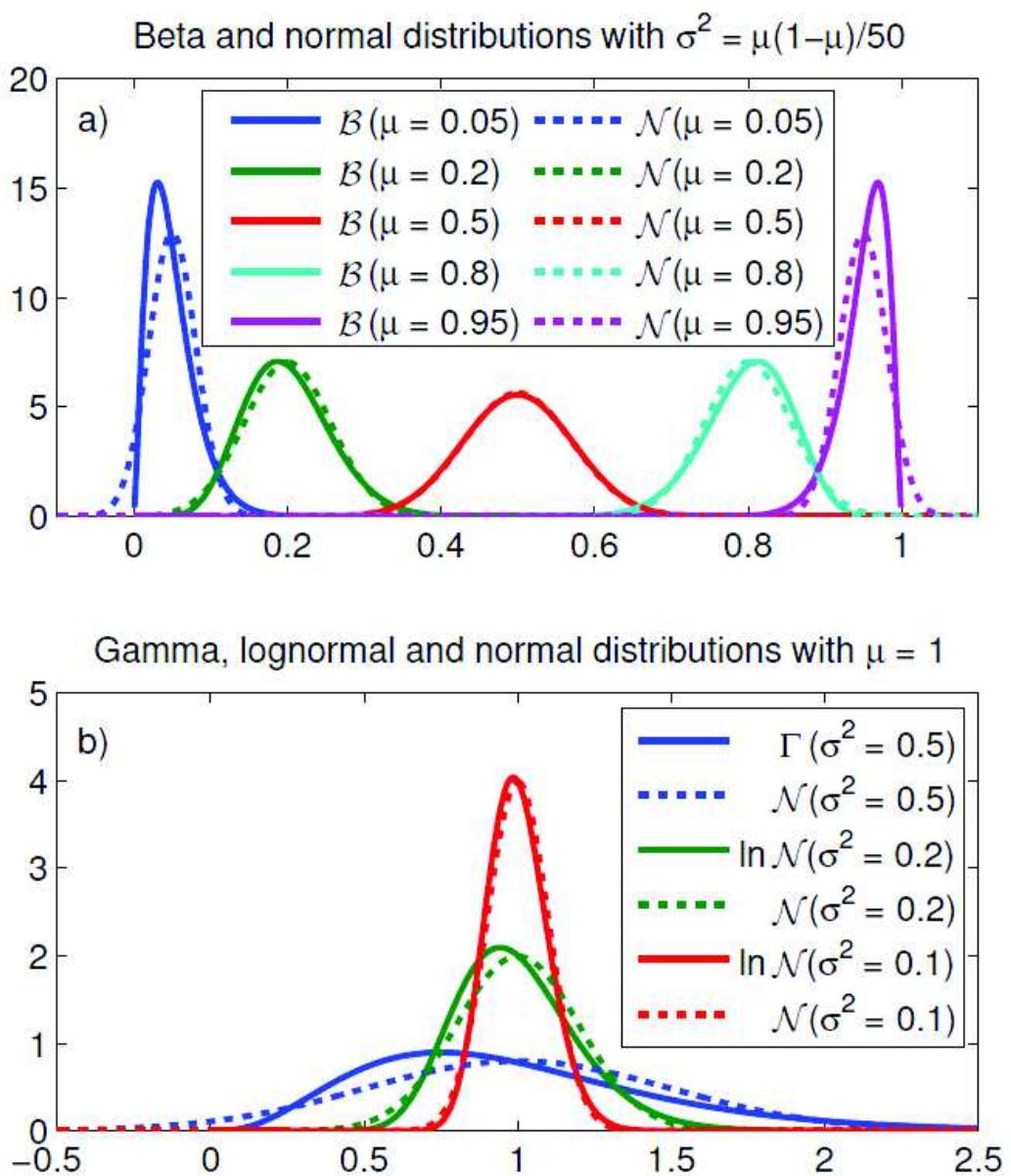
Figure R1. Examples of a) beta and b) gamma and lognormal distribution compared with their analogous normal distributions using the same mean and variance.

2) You need to provide a better justification to the use of these distributions on page 7.

AR: A paragraph has been added in the revised manuscript to justify the use of non-normal distributions. Essentially, it is to satisfy hard boundaries on observations and meteorological input. Since all observations and meteorological input in the study have limits (ex: range between [0 infinity] or [0 1]), adding a normally distributed perturbation can mean those limits are sometimes exceeded. A simple

way to get around this problem might be to set all exceeding values at the boundary (ex : all negative values set to zero). However, this introduces a bias, which is another and likely bigger problem. Other distributions were therefore used to satisfy the physical limits, while keeping some visual similarities with a normal distribution if possible.

3) You need to rewrite the paragraph starting on page 6 at line 14 as it is confusing as it would appear that it looks like you are referring to equations.

AR: As suggested by another Referee, the word "step" has been added before each number in parentheses to avoid confusion.

4) The statement on page 17, line 30 does not make sense and is confusing about the need for linear relationships which you really should have with the EnKF.

AR: The sentence is a relic of a previous formulation and has been removed.

5) On page 9 you are finishing the details about the localization but i am concerned that because you achieve this wrt the true state that this may not be the case in the real data situation and you need some sort of disclaimer here as you are kind of using the true localization which would not be the case in reality.

AR: This was meant to be discussed in the conclusion, in the paragraph explaining the assumptions and limitations of the results. This seems like a more appropriate section in the manuscript than the experimental design since 1) the results are not known yet in the experimental design section and 2) it allows us to generalize all results as valid within the synthetic limitations of the experiment. This avoids repetition since it applies to all steps, not only the covariance localization. An additional sentence has been added in the conclusion to explicitly state that dependency.

Minor comments:

1) Page 3, line 27 remove to

AR: This has been corrected in the revised manuscript.

2) Page 5, line 17, you mention the gain yet you have not defined it.

AR: The whole paragraph has been moved to the end of the section, as well as partly merged with the (previously) last paragraph and reworked to avoid repetition.

3) Page 11, line 33, remove the first that

AR: This has been corrected in the revised manuscript.

4) Page 15, line 15, sits is not a very scientific way to describe where the site is.

AR: The expression "sits at 682 mm" does not refer to the physical location of the site, but mean annual maximum SWE as described earlier in the sentence. Nonetheless, this has been rephrased to "contains a mean annual maximum SWE of 682 mm"

5) General comment. you use both Gaussian and normal please be consistent and only use one of them

AR: This has been corrected in the revised manuscript by using "normal" throughout the manuscript.

Comments to Referee #3 (Kevin He)

The authors present a set of synthetic experiments in assessing the potential added value of assimilating streamflow, SWE, SCA (via EnKF) into the CEQUEAU model in short- to medium-range streamflow forecasting at the Nechako watershed located in BC, Canada. Results indicate that streamflow DA and SWE DA lead to improvements in short-term forecast and medium-term forecast (during snow melt period), respectively. Assimilation of streamflow and SWE simultaneously yields even better results at both scales. However, assimilating SCA does not show any benefit. Overall, the methodology and results are sound and meaningful, yet not innovative. The paper is very well written and organized. I think it will be of interest to the readership of HESS.

My major comment is that, from the perspective of water resources management, the bias of the mean (or median) ESP forecast is typically an important factor considered in water-related decision making (e.g., water supply allocation, reservoir release/hydropower generation schedule, among others). In light of this, when assessing EPS forecast skill, the bias is normally analyzed.

AR: We have been working closely with Rio Tinto and they, like many other water resources managers, also consider the bias to be an important metric (maybe even the most important), especially during the melt period. While this has been computed for all scenarios, as with many other metrics, we felt this actually added little to the discussion that did not justify the doubling of the number of figures for forecasts. This is because we have generated bias-free synthetic observations and meteorological input. This approach will likely always result in a nonzero bias due to the non-linearity of the hydrological model (e.g. two distributions of the same amount of rain can lead to different quantities of cumulated streamflow due to evapotranspiration, etc.) and the finite period used in the study, but it should ideally be very small.

In our case, average streamflow bias is less than 1 % for the open loop compared with the true state. Zooming on each year at a time, simulated and true cumulated streamflow difference oscillates around 5 % on average over the 10 years considered. This is mainly why the assimilation of SWE can lead to some improvement. If bias was always perfect, adding or removing snow would not lead to improvements. Although this bias value is improved in various ways with data assimilation, the window for improving bias remains small however. For the real world Nechako basin, average bias is estimated at around 20 %.

In current works that has not yet been published, biased precipitations are purposely added to test the robustness of the approach. Real data assimilation has also been evaluated. In both of those cases, bias becomes a central part of the results as it is significantly far from null to begin with (there is something to potentially improve upon). As it is currently however, we feel that adding figures of bias would contribute very little to the discussion due to the near-ideal framework used in the synthetic experiment.

In the case of this study, the score MSSS is applied in the sensitivity analysis part (Figures 4, 6 and 7) but not the forecast part (Figures 8-12). The relevant results should be added (either in tabular or graphic form) and discussed.

AR: The use of MSSS during the sensitivity analysis part and CRPSS during the forecast part was done to facilitate comparisons between other similar studies. The MSSS and CRPSS are relatively similar metrics used in different contexts. The mean square error (non-normalized version of the MSSS) is often used as a metric duration calibration or comparison between two curves, while the CRPS is mostly used to evaluate ESPs since it is adapted to ensembles. In the limit where an ensemble reaches 1 member, the CRPS simply becomes the mean absolute error. Although it is not exactly the same as the mean square error, the two carry much of the same information for non-ensemble curves. For ensembles, one should use the mean or median to compute the MSSS, which could yield very different results than the CRPSS since the latter is sensitive to the precision of the ensemble and not only its accuracy. However, in our case, the two respond in a very similar fashion as can been seen in Fig. R2 (below) in this response. Similar graphs can be generated for all comparisons made using the CRPSS, but we believe this would add very little to the discussion.
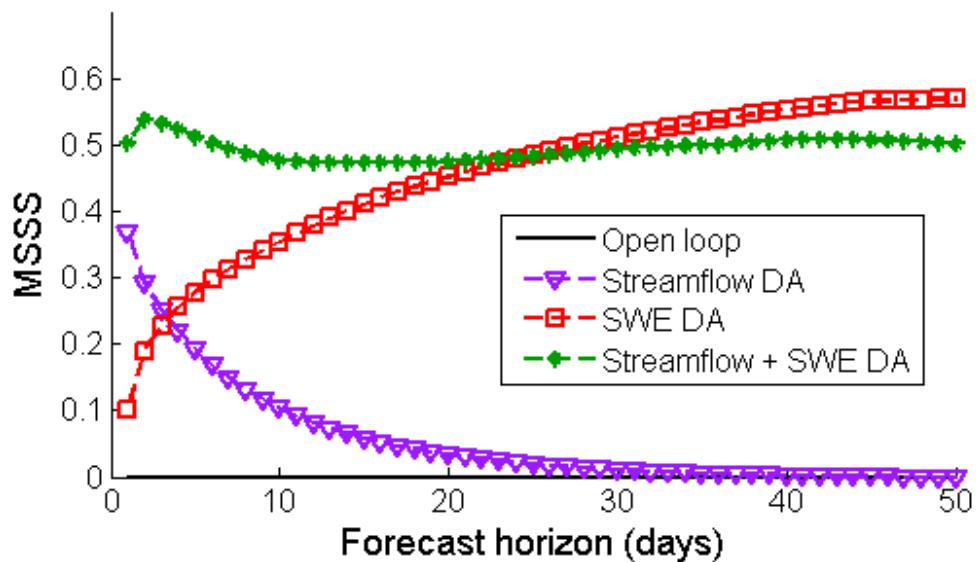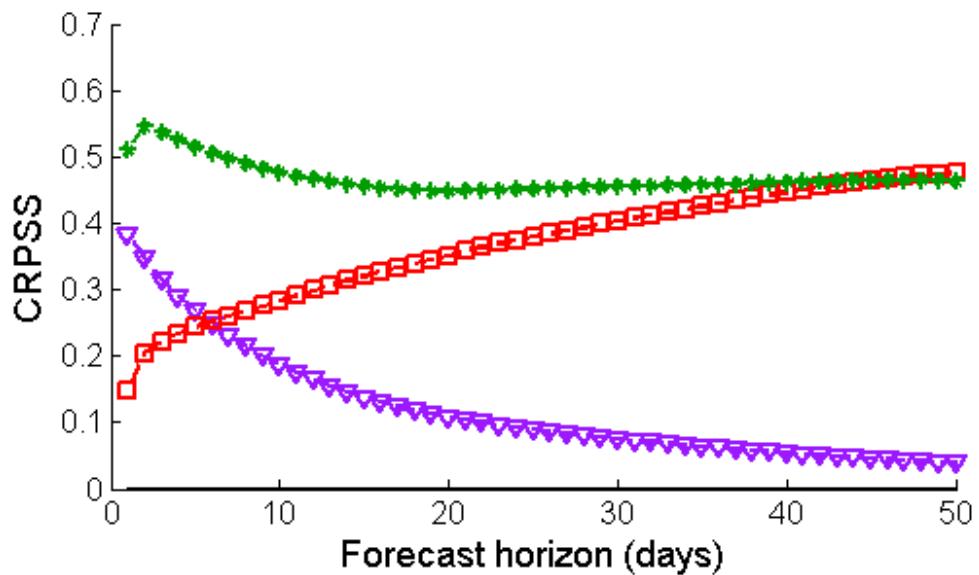
Figure R2. CRPSS and MSSS relative to the true state for a sample of forecasts.

My minor comments include 1) the authors need to be clear about how often the forecasts are issued (every day, once a week, or once per month in the study period from 8/15/1990 to 8/14/2000). If it is once a month, the authors need to discuss the sample size issue (10 years) when discussing the skill scores;

AR: This has been clarified in a new section (3.1.3) of the revised manuscript dedicated to ensemble streamflow predictions.

2) Line 7 of Page 2, "Franz" should be "Franz et al.";

AR: This has been corrected in the revised manuscript.

3) Lines 26-27 of Page 3, august should be August; delete "to";

AR: This has been corrected in the revised manuscript.

4) Line 8 of Page 4, (Fig. 2), Army Corps of Engineers;

AR: This has been corrected in the revised manuscript.

5) Lines 14-23 of Page 6, other than use (1), (2),. . ., it would less confusing when using (#1), (#2), . . ., or (Step 1), (Step 2), etc. ;

AR: This has been modified in the revised manuscript by using the term "step" before each number.

6) Line 5 of Page 7, delete "to";

AR: This is a typo and has been replaced with "the".

7) Line 6 of Page 8, modify "than";

AR: This has been corrected in the revised manuscript by removing that part of the sentence. This was done in rearrangement of the sections to split the information relating to the generation of synthetic observations and observation ensembles.

8) Lines 20-22 of Page 8, rework on the sentence;

AR: The sentence has been reworked in the revised manuscript to:

"This approach avoids introducing a systematic bias when assimilating SCA values at 0 or 100 %. When values are at 0 % (or 100 %), perturbations can only introduce higher (or lower) values in order to remain within the physical limits of the observations. This approach also gives the observations a greater uncertainty during the transition periods when SCA, which loosely follows the greater uncertainty attributed to MODIS observations over the same period (Hall and Riggs, 2007)."

9) Line 12 of Page 10, change MSS to MSSS;

AR: This has been corrected in the revised manuscript.

10) Line 20 of Page 10, in order to.

AR: This has been corrected in the revised manuscript.

Combined assimilation of streamflow and snow water equivalent for mid-term ensemble streamflow forecasts in snow-dominated regions

Jean M. Bergeron[1], Mélanie Trudel[1], Robert Leconte[1]

5 [1]Department of civil engineering, Université de Sherbrooke, Sherbrooke, J1K 2R1, Canada

Correspondence to: Jean M. Bergeron (Jean.Bergeron2@USherbrooke.ca)

Abstract. The potential of data assimilation for hydrologic predictions has been demonstrated in many research studies. Watersheds over which multiple observation types are available can potentially further benefit from data assimilation by having multiple updated states from which hydrologic predictions can be generated. However, the magnitude and time span

10 of the impact of the assimilation of an observation varies according not only to its type, but also to the variables included in the state vector. This study examines the impact of multivariate synthetic data assimilation using the Ensemble Kalman Filter (EnKF) into the spatially distributed hydrologic model CEQUEAU for the mountainous Nechako River located in British-Columbia, Canada. Synthetic data includes daily snow cover area (SCA), daily measurements of snow water equivalent (SWE) at three different locations and daily streamflow data at the watershed outlet. Results show a large variability of the

15 continuous rank probability skill score over a wide range of prediction horizons (days to weeks) depending on the state vector configuration and the type of observations assimilated. Overall, the variables most closely linearly linked to the observations are the ones worth considering adding to the state vector. The performance of the assimilation of basin-wide SCA, which does not have a decent proxy among potential state variables, does not surpass the open loop for any of the simulated variables. However, the assimilation of streamflow offers major improvements steadily throughout the year, but

20 mainly over the short-term (up to 5 days) forecast horizons, while the impact of the assimilation of SWE gains more importance during the snowmelt period over the mid-term (up to 50 days) forecast horizon compared with open loop. The combined assimilation of streamflow and SWE performs better than their individual counterparts, offering improvements over all forecast horizons considered and throughout the whole year, including the critical period of snowmelt. This highlights the potential benefit of using multivariate data assimilation for streamflow predictions in snow-dominated regions.

25 **Keywords**. Ensemble Kalman Filter, hydrological modeling, ensemble forecast, snow accumulation and melt, multivariate data assimilation

1 Introduction

Water resource management for reservoirs located in snow dominated regions relies on an accurate portrayal of the snow water equivalent (SWE) spatial and temporal distribution in order to make accurate streamflow predictions. Some water resources managers make use of Ensemble Streamflow Prediction (ESP) to plan reservoir operations over various lengths of time. ESPs have the benefit of integrating weather forecast uncertainty, either by making use of weather ensemble predictions (de Roo et al, 2003) or by using historical weather data (Day, 1985) as input in a hydrologic model. However, ESPs depend heavily on the model's initial conditions (Franz et al, 2008). Presently, many water resources managers still use a manual approach to adjust the initial state of the watershed based on available observations and the user's experience (Liu et al, 2012).

Data assimilation (DA) methods, such as the Ensemble Kalman Filter (EnKF; Evensen, 2003) can improve the estimation of the initial state of the watershed while also providing an uncertainty on this initial state (Liu and Gupta, 2007). Several authors have already shown the added value of DA in snow-dominated watersheds to improve the estimation of the state of the watershed (De Lannoy et al, 2012; Dechant and Moradkhani, 2011; Nagler et al, 2008; Slater and Clark, 2006; Andreadis and Lettenmaier, 2006). Some studies have also integrated DA in ensemble forecast systems for relatively short-term (up to 5-10 days) hydrologic forecasts (Abaza et al, 2015; Abaza et al, 2014; He et al., 2012), but studies focusing on longer forecast periods are scarce even though the need exists for water resource managers.

Multivariate DA applications in hydrology are becoming more frequent, but generally focus on streamflow and soil moisture (Samuel et al, 2014; Trudel et al, 2014; Lee et al, 2011), omitting snow water equivalent. In snow-dominated watersheds, the key initial states include not only information about the hydric state, such as soil moisture and streamflow, but also the snow cover state, such as snow water equivalent (SWE) and snow cover area (SCA). To the authors' knowledge, no published studies pertain to the combined assimilation of information about a watershed's hydric and snow state. Since the lasting impact of hydric DA and snow DA can be quite different given the different physical processes driving them, the simultaneous DA of both types of data could yield improvements over a potentially longer length of time.

However, data assimilation performance depends on various factors, such as the choice of variables to be updated by an observation (hereby referred to as the state vector configuration). Abaza et al. (2015) have demonstrated this importance when assimilating streamflow in a hydrologic model. Going from univariate to multivariate DA increases the number of degrees of freedom, which increases the complexity of the matter. ~~Is the state vector configuration still important in multivariate DA?~~The importance of state vector configuration when using multivariate DA for hydrological modeling has yet to be investigated.

2

The study's main objectives are to 1) investigate the potential impact that multivariate data assimilation of hydric (streamflow) and snow-related (SWE and SCA) data can have on short-term (1-5 days) and mid-term (up to 50 days) streamflow forecast, and 2) to explore how this impact varies as a function of the state vector configuration.

2 Materials and methods

2.1 Study area description and data

Simulations were conducted in a synthetic setting based on the Nechako watershed located in British-Columbia, Canada (Fig. 1). The watershed includes a reservoir, which drains an area of approximately 14000 km^2. The reservoir is managed by Rio Tinto mainly for hydroelectricity production purposes. The watershed includes part of the Coast Mountains in the west region, such that the difference in elevation between the highest and lowest point in the watershed reaches about 1700 m. At these latitude and altitude, most (estimated at 53%) of the precipitation falls as snow.

There are various types of data gathered regularly over the watershed. First are seven weather stations managed by Rio Tinto, three of which measure daily precipitation and air temperature only (yellow circles). Three others also include snow pillows (red squares), which measure snow water equivalent. The northernmost snow pillow is located at Mount Wells, the southernmost at Mount Pondosy and the westernmost at Tahtsa Lake. Maximum seasonal SWE observations average 615, 853 and 1393 mm for the Mount Wells, Mount Pondosy and Tahtsa Lake snow pillows respectively. The distribution of snow on the ground follows a strong East-West gradient such that measurements at Tahtsa Lake typically yield much more snow that Mount Well and Mount Pondosy. The northernmost weather station (blue triangle) is located next to the spillway at Skins Lake and also takes hydrometric measurements. Historical daily water levels can then be converted into natural inflows by also taking into account spilled and turbined flow. Finally, daily snow cover area (SCA) data derived from the spaceborne sensor MODIS/Terra are also considered (Hall et al, 2002). Because of its spatial coverage and relatively high temporal resolution, remotely sensed snow data from MODIS have proven to be valuable in a number of hydrologic studies (Bergeron et al, 2014; Roy et al, 2010; Tang and Lettenmaier, 2010; Andreadis and Lettenmaier, 2006; Clark et al, 2006), including one applied to the Nechako watershed (Marcil et al, 2016).

The meteorological observations gathered over a period of 10 year (from 15th august August 1990 to 14th august August 2000) will bewere used to as a basis upon which a synthetic experiment (see below) will tested the added value of three types of synthetic observations (streamflow, SWE and SCA) for data assimilation purposes. The only datasets actually used is were the meteorological station data, as well as some streamflow observations (owned by Rio Tinto) and MODIS/Terra daily L3 snow cover data (Hall et al, 2002) for the initial model calibration performed prior to the synthetic experiment presented in this manuscript. All thewhile observation data are were created synthetically to mimic streamflow, SWE and SCA data

that could be measured or estimated using hydrometric, snow pillow and MODIS data, respectively. More details on the creation of synthetic observations and meteorological input in sections 3.1.1 and 3.1.2 respectively.

2.2 Model description

The hydrologic model used ~~is~~ was the spatially distributed, conceptual model CEQUEAU (Charbonneau et al, 1977). It is currently being used by Rio Tinto ~~Alcan~~ to model hydrologic processes including streamflow at the outlet of the Nechako watershed, considered to be the spillway where the hydrometric station is also located. All variables are computed at a daily time step using a set of parameters to calibrate and daily meteorological input consisting of mean air temperature and precipitation. The set of parameters used in this study was the result of a manual calibration performed by Rio Tinto by comparing the simulated streamflow at the outlet with the corresponding real streamflow observations. A summary of the main processes concerning the production and transfer functions is presented here to facilitate the understanding of the state variables used in this study.

CEQUEAU divides the watershed into regular square pixels called "whole squares" over which the production function is computed (Fig. 2). The current version of CEQUEAU uses the snow model presented by the U.S. ~~army~~ Army ~~corps~~ Corps of ~~engineers~~ Engineers (1956) to simulate most snow-related processes. The SWE is actually computed separately for forested and open areas, which have their own set of parameters, but is aggregated here as a weighted sum according to the proportion of forested and open areas within each whole squares. The only variable computed separately (i.e. outside from CEQUEAU) is SCA, which is computed using a depletion curve (Anderson 1973). The depletion curve used here follows Andreadis and Lettenmaier (2006), which uses a three parameter beta distribution:

$$\mathrm{SCA}_i = \mathrm{B}^{-1}\left(\frac{\mathrm{SWE}_i}{\min(\mathrm{SWE}_{\max,i}, SI)}\,\middle|\,\alpha_{SCA}, \beta_{SCA}\right), \tag{1}$$

where SCA_i is the resulting snow cover area over a whole square i, SWE_i is the simulated snow water equivalent over the same area, $\mathrm{SWE}_{\max,i}$ is the annual maximum snow water equivalent since the beginning of the accumulation period over the same area, SI represents the value of SWE above which it is assumed there is always 100% snow cover and α_{SCA} and β_{SCA} are shape parameters for the beta distribution itself. The calibration of those three parameters was conducted using the SCE-UA method (Duan et al, 1992) to minimize the root mean square difference between simulated snow cover area and MODIS/Terra daily L3 snow cover data (Hall et al, 2002) averaged within each whole square within the Nechako watershed. ~~These three parameters were previously calibrated to match MODIS/Terra daily L3 snow cover data (Hall et al, 2002).~~ It is important to note that SCA is computed as an output only and is therefore not considered to be a state variable since it has no impact on future simulations if its value is ~~tempered~~ tampered with.

4

CEQUEAU then uses three conceptual reservoirs to simulate various hydrologic processes from the available water resulting from rain or snow melt. There is an optional lake reservoir, an upper reservoir (called "soil moisture reservoir" in this study) and a lower reservoir (called "groundwater reservoir" in this study).

All in all, the state variables simulated over each whole square includes snow water equivalent (SWE), a snow ripening index (SRI), a snow temperature index (STI), the soil moisture level (SML), the groundwater level (GWL) and the lake water level (LWL) should there be one. There are 644 whole squares in the case of the Nechako watershed.

Each whole square is itself divided into "partial squares" according to the subpixel drainage divide. There are a total of 1082 partial squares in the case of the Nechako watershed. Available water from whole squares (whether from surface runoff or drained from the soil and/or groundwater reservoirs) is divided into these partial squares according to the fraction of whole square area drained by partial squares to form volumes (VOL). These volumes that are represent the total amount of water available for transferred from one partial square to the next. The actual amount of water transferred at a specified rate to formover a given period is called streamflows and is defined as follows :

$$SF_j = \frac{1}{\Delta t} \sum_k^{M_j} ext_k \cdot \text{VOL}_k, \tag{2}$$

where SF_j is the streamflow at partial square j, M_j is the number of partial squares directly upstream, ext_k is a transfer coefficient and Δt is the time step. VOL is therefore a state variable, but Sstreamflow, like SCA, is not considered to be a state variable since it has no impact on future simulations if its value is tempered tampered with.

2.3 Ensemble Kalman Filtering

The Ensemble Kalman Filter (EnKF) is a data assimilation method developed by (Evensen (1994). It is an approach often used in hydrology, mainly due to its ability to consider non-linearities in the model and its relative simplicity to implement. The EnKF is a sequential method, meaning it relies only on current observations to update state variables as opposed to non-sequential approaches such as smoothers (Evensen and van Leeuwen, 2000) and recursive methods (McMillan et al, 2013) methods.

However, the method has practical and theoretical limitations. Firstly, the EnKF relies on an ensemble representation of model and observation errors that are valid in the limit where ensemble sizes approach infinity. This is not feasible in practice, so a finite sample is used instead which aims to be sufficiently large such that sampling errors are negligible while ensuring that computational power and memory limitations are met. The method also makes use of model and observation covariance matrices to compute the gain during the updating process. These covariance matrices assume a linear relationship between variables. Inaccurate analyses may result for cases where this assumption does not hold. Despite these limitations,

The EnKF propagates an ensemble of model runs based on a Monte Carlo implementation to represent model errors. The

5 model covariance matrix (P_t^b) at a time t is computed from the state vector (x_t^b) holding the N ensemble members and their simulated variables; and the ensemble mean of the state vector $(\overline{x_t^b})$, therefore implicitly taking the model dynamics into consideration:

$$P_t^b = \frac{1}{N-1}\left(x_t^b - \overline{x_t^b}\right)\left(x_t^b - \overline{x_t^b}\right)^\top,$$ (3)

When an observation is available, it is perturbed to form an ensemble of observations that are used to update each ensemble

10 member. The updating step applies the Kalman gain (K_t), which is computed from observation (R_t) and model covariance matrices as well as an observation operator (H_t), which relates the model states to the observation:

$$K_t = P_t^b H_t^\top (H_t P_t^b H_t^\top + R_t)^{-1},$$ (4)

The Kalman gain acts as a weighted average between the observation and state vector to yield a post-filter analysis (x_t^a) computed as such:

15 $$x_t^a = x_t^b + K_t(y_t - H_t x_t^b),$$ (5)

The EnKF has practical and theoretical limitations. Firstly, the EnKF relies on an ensemble representation of model and observation errors that are valid in the limit where ensemble sizes approach infinity. This is not feasible in practice, so a finite sample is used instead which aims to be sufficiently large such that sampling errors are negligible while ensuring that computational power and memory limitations are met. The method also makes use of model and observation covariance

20 matrices to compute the gain during the updating process. These covariance matrices assume a linear relationship between variables. Inaccurate analyses may result for cases where this assumption does not hold. Despite these limitations, the EnKF's usefulness in hydrology has nonetheless been shown in numerous studies, which is why it is also used in this study. The EnKF also assumes normally distributed, bias-free and time-independent errors for both the model and the observations. Since these assumptions are not always met, which means that optimality is not guaranteed, a synthetic experiment is

25 recommended to test the applicability of the method to the specific case being studied.

The Ensemble Kalman filter aims to minimize the resulting analysis error under several assumptions such as having normally distributed, bias free and time-independent errors, as well as linear relationships between variable errors during the updating step. Though these assumptions are not always met, which means that optimality is not guaranteed, the value of the

30 EnKF for hydrologic applications has nevertheless been shown repeatedly (Abaza et al, 2015; Samuel et al, 2014; Trudel et al, 2014; Brocca et al, 2012; Forman et al, 2012; Xie and Zhang, 2010; Weerts and El Serafy, 2006).

6

3 Experimental design

3.1 Synthetic experiment

Synthetic experiments, like the ones done by Xie and Zhang (2010) or Weerts and El Serafy (2006), are testbeds used ~~A synthetic experiment is a useful approach~~ to test the robustness of a data assimilation method or to tune various hyper-parameters. This is because the true state is known since it is initially created from known inputs.

For the current study, interpolated meteorological ~~observations~~ input and a specific set of parameters were used to run CEQUEAU, the output of which was considered to be the true state (step 1) (see Fig. 3). Synthetic observations, which include daily streamflow, SWE and SCA (step 2), and meteorological input, which include daily mean air temperature and precipitation (step 3), were then obtained by applying a perturbation to the true state and true meteorological input respectively. This means that the observations sets described thus far are not directly used, but synthetically generated using known parameters and perturbation. The synthetic observation ensemble (step 4) and meteorological ensemble (step 5) were created by further perturbing the synthetic observations and meteorological input. Ideally, these meteorological and observation ensemble perturbations should reflect the true errors of the synthetic meteorological input and observations for an optimal analysis. An ensemble of hydrologic states (step 6) was then obtained by running CEQUEAU using the synthetic meteorological ensemble. The EnKF was then applied using both the model and observation ensembles to produce an analysis (step 7), which was used as an initial state to produce ensemble streamflow predictions (ESP; step 8) using the true meteorological input. Additional details pertaining to the procedure used in generating perturbations and ensembles are included in the following sections.

~~Using the true meteorological input implies that over a sufficiently large forecast horizon, every DA scenario considered in this study is likely to converge to the true state, but at different rates. By comparing the relative gains in performance over the ensemble with no data assimilation (open loop), one can then observe the length of time upon which DA impacts ESPs without having erroneous meteorological input affecting the results.~~

3.1.1 Synthetic observation perturbation

Three types of observations ~~a~~were considered; namely streamflow, snow water equivalent (SWE) and snow cover area (SCA). These synthetic observations were generated using a daily time step since their real-world counterparts are usually available on a daily basis.

In order to abide by EnKF assumptions, observations errors should ideally have a normal distribution. However, this is not practical due to the physical limits of the observations. For example, SWE observations cannot be negative and adding a normally distributed perturbation to SWE could result in some values being negative. Raising the negative values to zero or

7

above would introduce a bias. Therefore, other distributions that share similarities with a normal distribution, while ensuring that physical limits are respected, were used to generate synthetic observations.

Synthetic watershed-wide SCA were created using perturbations that follow a beta distribution since SCA is bounded between 0 and 1. SCA observations are expressed as $y_{t,j} \sim \mathcal{B}^{-1}(Q_{t,j}|\alpha_{t,j}, \beta_{t,j})$, where $Q_{t,j}$ is the cumulative probability of a temporally correlated normal random field with zero mean and unit variance at time t for observation j, and $\alpha_{t,j}$ and $\beta_{t,j}$ are positively valued shape parameters. The shape parameters may be expressed in terms of the mean $\mu_{t,j}$ and variance $\sigma_{t,j}^2$, but it must follow that $\sigma_{t,j}^2 < \mu_{t,j}(1 - \mu_{t,j})$. The variance was arbitrarily set to $\sigma_{t,j}^2 = \mu_{t,j}(1 - \mu_{t,j})/50$, such that the resulting shape parameters are $\alpha_{t,j} = 49\mu_{t,j}$ and $\beta_{t,j} = 49(1 - \mu_{t,j})$. Examples of beta distributions for different means using the same definition of variance as described above are shown in Fig. 4a. The distribution has a null variance and greater deviation from a normal distribution when the snow covers either 0 or 100 % of the watershed, as well as a variance greatest and resembling most a normal distribution at 50 % SCA. This approach avoids introducing a systematic bias when assimilating extreme values of SCA. When values are at 0 % (or 100 %), perturbations can only introduce higher (or lower) values in order to remain within the physical limits of the observations. This approach also gives the observations a greater uncertainty during the transition periods when SCA, which loosely follows the greater uncertainty attributed to MODIS observations over the same period (Hall and Riggs, 2007).

Synthetic streamflow and SWE observations were created ~~in a similar fashion than synthetic meteorological inputs, but~~ using perturbations that have a lognormal distribution since both observations are bounded to the left at 0 and are theoretically unbounded to the right. Observation values $y_{t,ij}$ can be expressed as $y_{t,j} \sim \ln\mathcal{N}^{-1}(Q_{t,j}|\mu_{t,j}, \sigma_{t,j}^2)$ ~~$y_{t,t} \sim L^{-1}(Q_{t,t}|\mu_y, \sigma_y^2)$, where $Q_{t,t}$ is the cumulative probability of a temporally correlated Gaussian random field with zero mean and unit variance~~. The true state was used for $\mu_{y t,j}$ for both streamflow and SWE, while the relative variance $\sigma_{y t,j}^2$ was set to 20 % and 10 % respectively. The error distributions using these parameters are shown in Fig. 4b. ~~As with the meteorological input error, t~~The exact value of these variances is arbitrary for feasibility purposes. However, since the conclusions of this study will likely be used to help setup real-world applications, the variances chosen should ideally be relatively similar to the error of their corresponding real observations. Since these real observation errors are not known, rough estimates are used.

The use of these distributions is a compromise between the normal distribution of observations required by the EnKF and the physical limits of the observations without introducing a bias.

3.1.2 Meteorological input perturbation

Both the true daily precipitation and temperature values were perturbed using a gamma distribution, which has the benefit of generating positive values exclusively. Perturbations were implemented such that the ~~weather~~meteorological input $z_{t,i}$ (precipitation or temperature) at time t over the ~~partial~~whole square i is the result of the inverse gamma function given the cumulative probability $P_{t,i}$ of a spatially and temporally correlated ~~Gaussian~~normal random field with zero mean and unit variance. This can be expressed mathematically as $z_{t,i} \sim \Gamma^{-1}(P_{t,i}|\kappa_{t,i}\theta_{t,i})$, where $\kappa_{t,i}$ and $\theta_{t,i}$ are shape and scale factor respectively. The shape and scale factors can be expressed in terms of mean $\mu_{t,i}$ and variance $\sigma_{t,i}^2$, such that $\kappa_{t,i} = \mu_{t,i}^2/\sigma_{t,i}^2$ and $\theta_{t,i} = \sigma_{t,i}^2/\mu_{t,i}$. In this study, synthetic precipitations are generated using the value of the true precipitation for $\mu_{t,i}$ and a relative variance of 50%, such that $\sigma_{t,i}^2 = 0.5 \cdot \mu_{t,i}$. Figure 4 shows the resulting error distribution using these parameters. Similarly, perturbed temperatures use the true temperatures for $\mu_{t,i}$ and ~~an absolute~~ standard deviation of 1°C. Within the synthetic study where the feasibility of the approach is tested, the exact value of these perturbations is arbitrary, so long as it is coherent between scenarios. The values used were such that the ~~performance~~Nash-Sutcliffe efficiency (Nash and Sutcliffe, 1970) of the simulated streamflow resulting from CEQUEAU using the perturbed meteorological input compared with synthetic streamflow was roughly similar to the performance of the simulated streamflow using real-world meteorological input compared with real streamflow observations.~~mean streamflow ensemble generated using CEQUEAU was roughly similar to the performance of the simulated streamflow using real-world meteorological input compared with observations.~~

3.1.3 Ensemble streamflow predictions generation

ESPs were generated using the ensemble of state variables resulting from the EnKF as initial states, true meteorological input and true model parameters. Using the true meteorological input implies that over a sufficiently large forecast horizon, every DA scenario considered in this study is likely to converge to the true state, but at different rates. By comparing the relative gains in performance over the ensemble with no data assimilation (open loop), one can then observe the length of time upon which DA impacts ESPs without having erroneous meteorological input affecting the results. This still generates an ensemble of streamflows since each ensemble member has its own initial states (VOL, SWE, SRI, STI, SWL, GWL, LWL).

ESPs were generated everyday over the entire study period (10 years) using a forecast horizon spanning 50 days.

3.2 Hyper-parameter tuning

The use of the EnKF requires the tuning of hyper-parameters, such as model and observation errors, and ensemble size. Improper specification of these hyper-parameters could lead to filter divergence (Houtekamer and Mitchell, 1998).

9

3.2.1 Ensemble size

The ensemble size should ideally approach infinity to reduce ~~to~~ the impact of sampling when covariance matrices are computed, but this is not feasible given the limits of computing power and memory. In practice, the ensemble size is chosen such that computing time is more reasonable while ensuring that sampling error remains small.

5

Tests were carried out using ensemble sizes of 8, 16, 32, 64 and 128 members. An ensemble size of 64 members was used for this study. This number was chosen as a function of the stability between successive runs and computing resources available. It was found to be a reasonable trade-off between having sufficiently consistent results between simulations, such that the sampling error would be dwarfed in comparison with the impact of the actual data assimilation, without exceeding

10   the computing resources available.

3.2.2 Meteorological ~~input perturbation~~ensemble generation

~~In order to reduce the impact of model error misrepresentation, the p~~Perturbation factors similar to the ones used to generate synthetic meteorological inputs were used to generate an ensemble spread ~~were the same as the ones used to generate synthetic meteorological inputs~~. This means that at every time step, meteorological ensemble ($z'_{t,i}$) were generated using an

15   inverse gamma function given the cumulative probability $P'_{t,i}$ of a spatially and temporally correlated normal random field with zero mean and unit variance, mathematically expressed as $z'_{t,i} \sim \Gamma^{-1}\left(P'_{t,i}|\kappa'_{t,i}, \theta'_{t,i}\right)$, where the shape factors are defined by $\kappa'_{t,i} = \mu'^2_{t,i}/\sigma'^2_{t,i}$ and $\theta'_{t,i} = \sigma'^2_{t,i}/\mu'_{t,i}$. The prime symbol is used to distinguish between the ensemble variables/parameters and the synthetic variables/parameters. Precipitations were generated using the value of the synthetic precipitation ($z_{t,i}$) for $\mu'_{t,i}$ and a relative variance of 50%, such that $\sigma'^2_{t,i} = 0.5 \cdot z_{t,i}$, while temperature ensembles were

20   generated using synthetic temperatures for $\mu'_{t,i}$ and a standard deviation of 1°C. Using similar perturbation factors between synthetic and ensemble versions of the meteorological input reduced the probability of filter divergence cause by a misrepresentation of the model error. Errors from CEQUEAU-specific parameters were not taken into consideration, such that the parameter set used for the generation of the true state were the same for the ensemble generation.

25   ~~Both the daily precipitation and temperature values were perturbed using a gamma distribution, which has the benefit of generating positive values exclusively. Perturbations were implemented such that the weather input $z_{t,i}$ (precipitation or temperature) at time t over the partial square i is the result of the inverse gamma function given the cumulative probability $P_{t,i}$ of a spatially and temporally correlated Gaussian random field with zero mean and unit variance. This can be expressed mathematically as $z_{t,i} \sim \Gamma^{-1}\left(P_{t,i}|\kappa, \theta\right)$, where κ and θ are shape and scale factor respectively. The shape and scale factors can~~

30   ~~be expressed in terms of mean μ_z and variance σ_z^2, such that $\kappa = \mu_z^2/\sigma_z^2$ and $\theta = \sigma_z^2/\mu_z$. In this study, synthetic precipitations are generated using the value of the true precipitation for μ_z and a relative variance σ_z^2 of 50%. Similarly,~~

10

~~perturbed temperatures use the true temperatures for μ_z and an absolute standard deviation of 1°C. Within the synthetic study where the feasibility of the approach is tested, the exact value of these perturbations is arbitrary, so long as it is coherent between scenarios. The values used were such that the performance of the mean streamflow ensemble generated using CEQUEAU was roughly similar to the performance of the simulated streamflow using real-world meteorological input compared with observations.~~

3.2.3 ~~Synthetic observation error representation~~Observation ensemble generation

As with model error representation, the perturbation factors used to generate an ensemble of observations were ~~the same as~~similar to the ones used to generate synthetic observations. Streamflow and SWE observation ensembles were created using perturbations that have a lognormal distribution centered around the synthetic observations $z_{t,j}$ with a relative variance of 20 % and 10 % respectively. Watershed-wide SCA ensembles were created using a beta distribution centered around the synthetic observation $z_{t,j}$ with a variance of $\sigma'^2_{t,j} = z_{t,j}(1 - z_{t,j})/50$. Using similar perturbation factors avoided problems caused by a misrepresentation of the observation errors. ~~Three types of observations are considered; namely streamflow, snow water equivalent (SWE) and snow cover area (SCA).~~

~~Synthetic streamflow and SWE observations were created in a similar fashion than synthetic meteorological inputs, but using perturbations that have a lognormal distribution. Observations $y_{t,t}$ can be expressed as $y_{t,t} \sim L^{-1}(Q_{t,t}|\mu_y, \sigma_y^2)$, where $Q_{t,t}$ is the cumulative probability of a temporally correlated Gaussian random field with zero mean and unit variance. The true state was used for μ_y for both streamflow and SWE, while the relative variance σ_y^2 was set to 20 % and 10 % respectively. As with the meteorological input error, the exact value of these variances is arbitrary for feasibility purposes. However, since the conclusions of this study will likely be used to help setup real-world applications, the variances chosen should ideally be relatively similar to the error of their corresponding real observations. Since these real observation errors are not known, rough estimates are used.~~

~~Synthetic watershed-wide SCA were created using perturbations that follow a beta distribution since SCA is bounded between 0 and 1. SCA observations are expressed as $y_{t,t} \sim B^{-1}(Q_{t,t}|\alpha,\beta)$, where α and β are positively valued shape parameters. The shape parameters may be expressed in terms of μ_y and σ_y^2, but it must follow that $\sigma_y^2 < \mu_y(1 - \mu_y)$. The variance was arbitrarily set to $\sigma_y^2 = \mu_y(1 - \mu_y)/50$, such that the resulting shape parameters are $\alpha = 49\mu_y$ and $\beta = 49(1 - \mu_y)$. This translates as a null variance when the snow covers either 0 or 100 % of the watershed and a variance greatest at 50 % SCA. This prevented introducing a systematic bias when assimilating SCA values at 0 or 100 %, as well as giving the observations a greater uncertainty during the transition periods when SCA, which loosely follows the greater uncertainty attributed to MODIS observations over the same period (Hall and Riggs, 2007).~~

11

3.2.4 Covariance localization

The main disadvantage in using a finite sample to compute covariance matrices is that the resulting covariance matrices are not exact. This may result in theoretically zero covariance elements between two theoretically uncorrelated variables to become small, but non-zero, which may deteriorate the performance of the EnKF.

5

One way to overcome this dilemma is to use covariance localization, where covariances are forced to zero between some variables. One option is the Schur product where a covariance matrix is multiplied element-wise by a distance-dependant correlation function (Houtekamer and Mitchell, 2001). However, there are other geophysical characteristics, such as landcover and elevation, which could be considered in the covariance localization. This would further increase the number of

10 parameters to set and the degree of subjectivity in setting those parameters when the degrees of dependence are unknown.

Another approach was used in this study, which is based on the improvements observed in the state vector. First, the open loop is executed, as well as a data assimilation scenario with one observation and the corresponding spatialized state variable included in the state vector (ex: 1 snow pillow assimilated and all modelled SWE included in the state vector). Then, the two

15 runs are compared with the true state on a spatial basis. In the case of CEQUEAU, these can be whole or partial squares depending on the variable analysed. The covariance matrix is localized such that the areas that do not show an improvement for the data assimilation scenario over the open loop are set to zero. This process is repeated for each observation.

While this process remains susceptible to the sampling error from the finite ensemble size, it is a simple approach that

20 exploits the availability of the true state in a synthetic experiment and limits the state vector size according to observed improvements.

In this study, only SWE observations have a corresponding state variable, so covariance localization ~~will~~ has only be~~en~~ applied to the SWE variable.

25 ### 3.2.5 State vector configuration

Though the state vector often comprises only of the variables corresponding to the observations or those judged to be relevant enough by the user, there are potentially many state variables which could benefit from the assimilation of available data if there exists a linear (or approximately linear) relationship between the modelled variables and the observations.

30 To determine which variable could benefit from being included in the state vector, one could execute multiple scenarios where each possible combination is compared with the true state. However, this could get very laborious even for a relatively small number of state variables. The current approach suggests reducing this number by first adding state variables one at a

time. The variables which show a global improvement can then be added to the state vector. Assuming that not all variables are added to the state vector, this reduces the number of combinations to try.

3.3 Metrics

Various metrics ~~are~~ were used to quantify results. The Mean Square Skill Score (MSSS), based on the Mean Square Error (MSE)~~,~~ ~~are~~ was used to assess the differences between various data assimilation scenarios and the open loop during the state vector configuration and covariance localization processes. The MSE for a variable of interest x is defined as:

$$\text{MSE}(x) = \frac{1}{N}\sum_{t=1}^{N}(\overline{x_t} - x_t^{\text{T}})^2, \tag{6}$$

where N is the number of time steps, $\overline{x_t}$ is the ensemble mean analysis of the state variable of interest at time t and x_t^{T} is the corresponding true state. It is often more convenient to express this score as a unitless skill score:

$$\text{MSSS}(x) = 1 - \frac{\text{MSE}(x)}{\text{MSE}_{\text{ref}}(x)}, \tag{7}$$

where $\text{MSE}_{\text{ref}}(x)$ is a mean square error of reference; the open loop in this case. The MSSS is bounded by $[-\infty, 1]$ and indicates an improvement as the skill score increases. Values above zero indicate an improvement over the reference (open loop) and a value of one indicates a perfect score; a perfect correspondence between the mean of the analysis and the true state.

The ensemble forecast performance ~~is~~ was assessed using the Continuous Rank Probability Score (CRPS; Hersbach, 2000) and its associated skill score (CRPSS). For this synthetic study, the CRPS is adapted as follows:

$$\text{CRPS}(x, f) = \frac{1}{N}\sum_{t=1}^{N}\int_{-\infty}^{+\infty}\left(F(x_t^f) - F(x_t^{\text{T}})\right)^2 dx, \tag{8}$$

where $F(x_t^f)$ and $F(x_t^{\text{T}})$ are the cumulative distribution function of the ensemble forecast at a horizon f and the true state, respectively. The CRPS has the same units as the variable of interest and is bounded by $[0, +\infty]$. A lower CRPS is a better score. As with the MSE and MSS<u>S</u>, it is often convenient to express the CRPS in its skill score form:

$$\text{CRPSS}(x, f) = 1 - \frac{\text{CRPS}(x, f)}{\text{CRPS}_{\text{ref}}(x, f)}, \tag{9}$$

where $\text{CRPS}_{\text{ref}}(x, f)$ is the continuous rank probability score of the open loop used as a reference in this case. Like the MSSS, the CRPSS is bounded by $[-\infty, 1]$, with higher values indicating a better score. Values above zero indicate an improvement over the reference and a value of one indicates a perfect score.

4 Results and discussion

4.1 State vector configuration and covariance localization

Before investigating the effect of data assimilation on streamflow forecasts, a state vector configuration analysis is̶ was conducted. This is̶ was done in order to find out which variables, among the 7 listed previously (VOL, SWE, SRI, STI, SWL,

5 GWL, LWL), should be included in the state vector for each type of data assimilated in order to reduce the number of comparisons to make.

4.1.1 Streamflow data assimilation

First presented are the results from the case where only streamflow at the outlet is̶ was assimilated. Streamflow at the outlet is computed by the model, but it is output-only. Therefore, in order for the assimilation of streamflow to have any impact on

10 the modelled states, additional variables must̶ needed to be added to the state vector.

Figure 4̶ 5 shows a boxplot of the MSSS computed for each variable on the whole watershed when they are individually included in the state vector using the open loop scenario as a reference. Values above 0 means there is an improvement for a particular partial (for volumes) or whole (for other state variables) square compared with the open loop. The boxes range

15 between the 25th and 75th percentiles, with a red bar to show the median, and the whiskers range between the maximum and minimum values. Outliers are not shown for visibility purposes. Results for the case where water volumes (VOL) are included along with the streamflow at the outlet show an improved score for each partial square on the watershed. This is not entirely surprising given the close relationship between streamflow and volume. This suggests a necessity to include VOL in the state vector when assimilating streamflow at the outlet for streamflow predictions.

20

Results also show a deterioration of snow water equivalent (SWE) for nearly 75% of whole squares on the watershed. Although there is some improvement for some whole squares, this suggests that including SWE in the state vector when assimilating streamflow at the outlet is unlikely to be beneficial for streamflow predictions. Although SWE does have an important impact on streamflow, there is a time lag between the snowmelt occurrence and the increase of streamflow at the

25 outlet. Since the EnKF assumes linear relationships between variables, the non-linearity between SWE and streamflow can result in a non-optimal analysis. In this case, the results are actually worse than open loop for most whole squares. Clark et al,̶ . (2008) discuss the issue of non-linearities between streamflow and other variables. To overcome this issue, one could use either a recursive approach, which allow adjustments of previously simulated variables, or smoother approach to DA, which also uses "future" observations to update current state variables. However, this may not be necessary given the

30 positive impact of streamflow DA on VOL, as well as in a multivariate DA scenario where other variables, such as SWE in this case, are also assimilated.

14

As for the snow ripening index (SRI) and snow temperature index (STI), the median sits around 0, meaning there is no improvement for 50% of the whole squares. This suggests little change can be obtained in the analysis by including those variables in the state vector. Similar to the case with SWE, there is likely a time lag issue between streamflow and these variables. However, there is also a weaker link between these variables such that a change in SRI or STI is not as strongly linked to an eventual change in streamflow as much as it is for a change in SWE.

Finally, results for the three conceptual reservoirs soil moisture level (SML), groundwater level (GWL) and lake water level (LWL) show an improvement for over half of the whole squares, with a greater number of whole squares improved for the GWL and slightly above 0 median for SML. This suggests that including these three variables in the state vector can potentially yield improvements for streamflow predictions. Though the relationship between the water level in these conceptual reservoirs and streamflow at the outlet is not exactly linear, mainly due to reservoirs having multiple orifices (see Fig. 2) and the time lag before water reaches the outlet, it may be sufficiently near linear ~~that~~ such that streamflow DA yields an overall improvement for most whole squares. For example, the median correlation coefficient of a simple linear regression between each reservoir for each whole square and the streamflow at the outlet is 0.12, 0.49 and 0.15 for SML, GWL and LWL respectively.

Samuel et al (2014) and Trudel et al (2014) found that updating soil moisture with streamflow observations actually deteriorated soil moisture simulation compared with real soil moisture observations. However, there are notable differences between these studies and the present one. Aside from the different features of the study area and model structure, the use of synthetic data instead of real data likely strengthens the link between variables and observations. Since synthetic observations are constructed using the same model and parameters as the model in which the observations are assimilated, there is no difference in scale between observations and modelled variables, which is often an important source of error for studies using real data.

Nonetheless, ~~T~~the inclusion of VOL, SML, GWL and LWL in the state vector were ~~will therefore be~~ considered during the assimilation of streamflow at the outlet. The impact of each scenario for streamflow predictions ~~will be~~are compared in section 4.2.1.

4.1.2 SWE data assimilation

The same analysis ~~is~~ was performed for SWE data assimilation from synthetic snow pillows. However, unlike streamflow, SWE is a state variable such that any changes made upon it will have repercussions on future simulations. SWE at the location of the snow pillows should therefore be included in the state vector and also potentially whole squares in the vicinity that are correlated with these locations. A spatial analysis ~~is~~ was performed first to determine the spatial extent upon which each snow pillow may affect modelled SWE in other whole squares.

Figure 5 6 shows the MSSS of SWE on the spatial level for the snow pillows located at Mount Wells, Mount Pondosy and Tahtsa Lake, using the open loop scenario as a reference. For each figure, the whole square which shows the most improvement is the area where the corresponding snow pillow is located. Whole squares which show improvements are mainly located around snow pillows, but the range differs for each snow pillows. Various areas in remote locations also show improvements for each snow pillow. As mentioned earlier, relationship with geophysical factors, such as distance from snow pillow, elevation and land cover, could be used to explain this variation, but a simpler approach was used such that the covariance localization was limited to whole squares showing improvements only. The covariance elements representing all the other whole squares were set to zero.

As for the state vector configuration, Fig. 6 7 shows the MSSS computed for each variable within the extent of whole squares which were positively impacted by SWE DA during the covariance localization process. The open loop scenario was used again as a reference. The results show no significant improvement for any other variable except for SWE itself, which yields only positive MSSS values by design. The lack of overall improvement for water-related variables (VOL, SML, GWL, LWL) is coherent with the time delay with changes in SWE. As for the other snow-related variables (STI, SRI), although there may be a relationship with SWE, it is non-linear (U.S. army Army corps Corps of engineersEngineers, 1956), which is further weakened by the distance separating SWE at a snow pillow from STI or SRI at another location.

Only the inclusion of SWE surrounding a given snow pillow in the state vector will beare considered during the assimilation of SWE for streamflow predictions in section 4.2.2.

4.1.3 SCA data assimilation

Like streamflow, snow cover area (SCA) is not a state variable. It is computed in parallel with CEQUEAU without having any direct effect on future simulations. In order to have any impact during the assimilation process, there must exist a linear or sufficiently near linear correlation between SCA and state variables. The update step should bring improvements to the state variables if the computed correlation also reflects the true correlation.

Figure 7 8 shows a boxplot of the MSSS computed for each variable when they are individually included in the state vector. The open loop scenario is used as a reference. Results show that global snow cover data assimilation yields little or no improvement for all state variables compared to the open loop scenario. For most cases, a strong deterioration is observed, suggesting that the current use ofway SCA data is used in this study is not well adapted for the current assimilation purposes using the EnKF. Marcil et al. (2016) have shown that there exists a relationship between the SCA and the percentage of cumulated streamflow at the outlet, but it is neither linear nor is cumulated streamflow a state variable. The EnKF

16

requirement that relationships between variables be linear and synchronized severely limits the value of global SCA data for the current application. This result is coherent with the findings presented by Clark et al (2006). Using a model which incorporates snow cover area as a state variable, such as the Snowmelt Runoff Model (SRM; Martinec, 1974) or the Soil and Water Assessment Tool (SWAT; Arnold et al, 1998), could overcome the issue of nonlinearities between variables, while using recursive or smoother approaches to data assimilation could help with the time lag issue between observations and state variables.

Given the absence of overall improvement for all the state variables, the impact of SCA DA on streamflow predictions ~~will~~ was not ~~be~~ considered in this study.

4.2 Streamflow forecasts

Aside from granting insight into the sensitivity of the system to the state vector configuration, the analysis in the previous section presented a list of state vector configurations likely to favour streamflow predictions improvements based on the improvement of various state variables. This section ~~will~~ present<u>s</u> <u>ensemble</u> streamflow prediction results for each configuration selected for each type of data assimilated.

4.2.1 Streamflow data assimilation

Focusing on the case where only streamflow at the outlet are assimilated, Fig. ~~8~~ <u>9</u> presents the CRPSS of predicted streamflow at the outlet over a forecast horizon of 50 days using the open loop as a reference. Only the state vector configurations that showed some improvements in the state vector configuration analysis section are shown.

Firstly, high values of CRPSS for short-term forecasts can be observed for the case where only volumes are included in the state vector (blue curve). The CRPSS subsides asymptotically to zero over time, which shows assimilating streamflow to update volumes improves streamflow predictions compared to the open loop only for a few days, after which the impact of streamflow assimilation becomes insignificant. <u>The duration of the impact depends on the residence time of the water stored in the model's partial squares (VOL). A relatively short-lived impact would mean a relatively short residence time. The high initial impact</u> ~~This~~ is not surprising given the nearly linear relationship between streamflows and volumes. <u>Assimilation of streamflow observations generate a globally positive update on volumes, as seen in</u> ~~This is coherent with the overall improvement observed for volumes in the state vector configuration analysis (~~Fig. <u>5</u>~~4).~~<u>,</u> <u>which in turn strongly affects simulated streamflows.</u>

Secondly, adding each of the three water reservoirs individually to the volumes yields different results. Even though lake water levels showed improvements over the majority of whole squares, the impact on streamflow predictions (red curve) is

17

marginal compared to the case where only volumes are included in the state vector. This is because the weights attributed to lakes in CEQUEAU are very low for most whole squares. Only about 0.5% of the entire watershed is modelled using the conceptual lake reservoir and its parameters, unlike the soil moisture and groundwater reservoirs, which are present in every whole squares. Adding SML (green curve) or GWL (magenta curve) instead of LWL increases not only the initial CRPSS, but also slows the decrease over time. This is consistent with the improvements observed for the updated water levels for over half of the whole squares compared with the open loop case, which translate as added improvements over the case where only volumes are included in the state vector. The slower decrease over time is also coherent with the increase in time it takes for water in the reservoirs to reach the outlet compared with water already in the routing system. The groundwater reservoir is shown to have an initially similar, but longer-lasting positive impact than the soil moisture reservoir. The soil moisture reservoir controls mainly the fast-flowing surface runoff, the amount of evapotranspiration leaving the system and the amount of water infiltrating into the ground–water reservoir. The ground–water reservoir has a numerically unlimited capacity, with no way out for the water except through evapotranspiration and the outlets that feed the routing system, making its impact on streamflows last longer than the relatively ephemeral soil moisture reservoir.

Finally, the scenario where all four variables are added to the state vector is analysed (orange curve). The difference noted with the other curves is mainly caused by the simultaneous inclusion of SML and GWL, since all the other curves already have VOL included, while LWL was already shown to have very little impact on streamflow at the outlet. Comparing the fully combined case with the VOL+GWL case, although the initial improvement is similar between the two, with the former slightly above, the latter has a slower decrease over time. This suggests that the addition of SML interferes with the GWL update. As seen for the state vector configuration analysis (Fig. 45), the assimilation of streamflow at the outlet had a positive impact on a greater number of whole squares for the GWL than the SML. Here, the increased number of deteriorated SML, which infiltrates into the groundwater reservoirs, hinders the GWL updates such that the results show some deterioration compared with the VOL+GWL case, even though it is still an improvement over the VOL only updates.

These results have some similarities and differences with other studies. For example, Abaza et al (2015) assimilated streamflow at the outlet using the EnKF to update two state variables (soil moisture in the intermediate and deep layers of the hydrological model used in their study) using a time step of 3 hours. The resulting gain in CRPS was high for the first time step and decreased quickly as a function of the forecast horizon such that mainly the first 24h benefited from the data assimilation. Chen et al (2013) found similar results with multiple performance criteria when assimilating streamflow using a variant of the EnKF (the ensemble square-root filter) to update various state variables represented by conceptual reservoirs. The observed improvement duration was even shorter, lasting less than 12 hours during flash flood events. The difference in impact duration is likely related to the different water retention time in each watershed. These studies were conducted on much smaller watersheds (all less than 800 km^2) than the Nechako watershed (around 14000 km^2), further highlighting that the performance of assimilation techniques is related to watershed characteristics.

18

4.2.2 SWE data assimilation

Following the same method as with streamflow, this section focuses on the case where only SWE from snow pillows ~~are~~ were assimilated. Since the state vector configuration analysis showed only improvement for the SWE variable, it ~~will be~~was the only variable added to the state vector for streamflow forecast. However, since there are three observations available, Fig. ~~9~~ 10 presents the CRPSS of predicted streamflow at the outlet when assimilating SWE from each snow pillow individually and collectively.

An interesting result is that the impact of each snow pillow on streamflow predictions varies greatly. The impact of the snow pillows located at Mount Wells (green curve) and Mount Pondosy (blue curve) are dwarfed in comparison with the impact of the snow pillow located at Tahtsa Lake (magenta curve). This is coherent with results from Marcil et al (2016) over the same watershed. The lower impact of the Mount Pondosy snow pillow is explained by the relatively small region of influence observed in Fig. ~~5b~~6b. As for Mount Well, even though it has the largest area of influence (Fig. ~~5a~~6a), it is also the snow pillow affecting the regions with the lowest altitudes and also the least amount of maximum SWE. Although the region~~s~~ affected by the Mount Wells snow pillow contain~~s~~ a mean annual maximum SWE of 410 mm, it is 40% less than for the region affected by the Tahtsa Lake snow pillow, which ~~sits at~~contains a mean annual maximum SWE of 682 mm.

Nonetheless, assimilating all three snow pillows yields better results for mid-term streamflow forecasts (Fig. ~~9~~10, red curve). Even though the Tahtsa Lake snow pillow carries the most importance, the other snow pillows have a positive effect on regions that are not reached by the Tahtsa Lake snow pillow area of effect. The assimilation of all three snow pillows does yield short-term forecasts improvements, but a better score is reached over time. This is because the impact of SWE over streamflow at the outlet occurs during snowmelt, which can occur at a much later date than when SWE observations are assimilated. Although the curve gives the impression of a monotonous increase over time, this is only due to the limit imposed on the forecast horizon. At a further horizon, the curve should eventually peak and decrease asymptotically to zero since there is no accumulation of snow from year to year. Over time, the simulation should eventually become indistinguishable from the open loop scenario.

Franz et al (2014) also assessed the impact of SWE data assimilation on ensemble predictions, but using real observations. Their results showed little improvement of forecast performance through SWE data assimilation, but they highlighted the role of a possible bias in the observations, as well as the difference in scale between the point-scale observations and the basin average SWE simulated by the model. In this synthetic experiment, no bias was specified on observations and there is no difference in scale between modelled and observed SWE, which could explain the differences observed between the two

19

studies. Biased observations and meteorological input were purposely omitted in this study, but may be added in future works to test the robustness of the approach.

4.2.3 Combined streamflow and SWE data assimilation

The focus now shifts to the case where streamflow at the outlet are simultaneously assimilated along with SWE from the three snow pillows. The state vector configuration which provided the best results from the streamflow data assimilation case are used (VOL+GWL) along with the best configuration from SWE data assimilation (SWE only). Although these configurations worked best with their respective data assimilation case, they could behave differently when both streamflow and SWE are assimilated together.

Table 1 presents four configurations for the combined assimilation of streamflow and SWE observations. These configurations differ in the overlap of their effect during the update phase such that some configurations allow both observations to simultaneously update the same variable, while others do not.

The performance of these configurations on the CRPSS for predicted streamflow at the outlet is presented in Fig. ~~10~~11. While all four configurations perform in a very similar way for short-term streamflow predictions, the group forms two pairs that differ in that the blue-magenta group allows SWE observations to update modelled streamflow, while the green-red pair does not. Although allowing SWE data assimilation to update VOL and GWL changes very little, a drop in performance occurs if streamflow assimilation updates modelled SWE. This is coherent with the state vector configuration analysis performed in the previous section (Fig. ~~4~~5 and Fig. ~~6~~7), where SWE data assimilation is shown to have a weak impact on VOL and a median MSSS around 0 for GWL, while streamflow assimilation deteriorated around 75% of SWE whole squares when they were included in the state vector.

Overall, the simultaneous assimilation of streamflow and observed SWE yields important improvements over the entire forecast horizon analysed, with the streamflow data assimilation improving mainly short-term streamflow forecasts and SWE data assimilation improving mainly mid-term streamflow forecasts. CRPSS values for combined assimilation of both streamflow and SWE observations were superior to CRPSS values for individual assimilation of streamflow or SWE over all forecast horizons, with the exception of forecast horizons higher than 45 days, where CRPSS values for SWE DA are slightly higher. This reveals that while the updated VOL and GWL by streamflow data assimilation may be very beneficial for short-term forecasts, they do not further improve the mid-term forecasts when combined with SWE data assimilation in comparison with the scenario where only SWE data is assimilated.

Furthermore, the assimilation of each~~the two~~ data types (streamflow and SWE) differs not only by their impact over the forecast horizon, but also over the time of the year. Fig. ~~11~~12 and Fig. ~~12~~13 show the monthly CRPS for the open loop

20

(black curve), the streamflow data assimilation including VOL and GWL (blue curve), the SWE data assimilation of all snow pillows (red curve) and the simultaneous, but separated, streamflow and SWE data assimilation (VG-S; green curve) for short-term (average of horizon from 1 to 5 days) and mid-term (average of horizon from 25 to 50 days) streamflow forecasts. The CRPS is shown for the open loop to show the performance change over the time of the year and the period when improvement is most needed. Recall that the CRPS ranges from zero to infinity, with zero being a perfect forecast. The period from May to July, which corresponds to the melt period, is therefore the period when the CRPS is highest for short- and mid-term forecasts are the most problematic. The scores for mid-term forecasts are lower than for short-term forecasts because the true weather is used as input for forecasts such that the open loop slowly converges to the true states over time.

Assimilating streamflow results in an improved score over the entire year for short-term forecasts, although little gain is obtained for mid-term forecasts. This steady improvement is to be expected since streamflow here is always nonzero and observations are available all-year round. On the other hand, the impact of the assimilation SWE from snow pillows is limited mainly to the melt period for both short-term and mid-term forecasts. However, this period corresponds to the problematic period when most gain can be obtained. The assimilation of SWE provides a better score than the assimilation of streamflow for the same period and improves going from short-term to mid-term forecasts. SWE assimilation complements streamflow assimilation as observed from the performance of the simultaneous assimilation, which yields both the steady improvements over the year and the important gain during the snowmelt period.

5 Conclusion

This study investigated the impact that multivariate data assimilation can have on streamflow forecasts using the CEQUEAU hydrologic model applied over the Nechako watershed in a synthetic experiment. The study also showed the importance of the state vector configuration on streamflow forecasts when using the EnKF.

Streamflow data assimilation was found to improve short-term streamflow forecast considerably. However, the impact dissipated relatively rapidly as a function of the forecast horizon, which was slowed by adding groundwater conceptual reservoir levels to the state vector. Improvements were observed for all months of the year; low-flow and high-flow periods alike.

On the other hand, the assimilation of snow water equivalent data from synthetic snow pillow data yielded streamflow forecast improvements mainly during the snowmelt period. Although the period lasts approximately three months, the impact was found to be greater than streamflow data assimilation over the same period. It was also noted that assimilating each snow pillow data individually yielded different results, with various radii of influence, such that the improvement from assimilating all three snow pillows simultaneously covered most of the watershed and yielded streamflow forecasts which

21

outperformed forecasts from any single snow pillow data assimilation. Over the forecast horizon, the peak of improvement was greater than or equal to the 50 days limit over which forecasts were simulated, which contrasts the short-lived impact of streamflow data assimilation.

5   Given their complementarity, streamflow and snow water equivalent data were assimilated simultaneously. The resulting streamflow forecast inherited the strengths from both types of data, having a strong, positive impact for both short-term and mid-term forecasts. Improvements were obtained for all periods of the year, but mainly during the snowmelt period, which is normally the most problematic.

10   The assimilation of basin-wide snow cover area failed to improve the simulation of any state variable, which deteriorated streamflow forecasts. The most probable factor was determined to be the absence of snow cover area as a state variable or a proxy with a sufficiently linear relationship with SCA. This was compounded by the nonlinear relationship between SCA and the state variables, which is required by the EnKF. Suggestions to improve the method to accommodate for snow cover area are to use a model which incorporates snow cover area as a state variable and/or to use a data assimilation approach 15   which takes into account a time lag between observations and state variables.

The results obtained are conditional to some hypothesesassumptions and limitations. First, all results depend on the general method and parameters used in creating the synthetic framework. sSince this is a synthetic experiment, it is assumed that a real experiment would behave similarly to a simulation using CEQUEAU with a specific set of parameters and inputs. 20   Second, the potential impact of data assimilation on streamflow forecasts observed depended on using the true weather inputs. Using real weather inputs may decrease this impact. Third, it is assumed that the error representations for the model inputs and the observations are known. In this study, they have been generated using specific distributions and variances to compromise between the need for normal distributions and the need to remain within the physical limits of the variables without introducing a bias, independent and unbiased. Finally, the impact of errors from the model parameters is assumed to 25   be negligible, such that the set of parameters was not altered from the true simulation's set of parameters.

The assumptions and limitationshypotheses listed reveal several challenges posed by the assimilation of multiple types of observations for streamflow forecasting purposes. Future works include investigating into the performance of multivariate data assimilation in the presence of biases and unknown errors, as well as the economic impact of streamflow forecasts 30   generated with multivariate data assimilation on real management practices.

22

respectively. We are also grateful for the two anonymous referees and Kevin He for taking the time to provide constructive comments.

References

5 Abaza, M., Anctil, F., Fortin, V. and Turcotte, R.: Sequential streamflow assimilation for short-term hydrological ensemble forecasting, J. Hydrol., 519, 2692–2706, doi:10.1016/j.jhydrol.2014.08.038, 2014.

Abaza, M., Anctil, F., Fortin, V. and Turcotte, R.: Exploration of sequential streamflow assimilation in snow dominated watersheds, Adv. Water Resour., 80, 79–89, doi:10.1016/j.advwatres.2015.03.011, 2015.

Anderson, E. A.: National Weather Service river forecast system: Snow accumulation and ablation model, 1973.

10 Andreadis, K. M. and Lettenmaier, D. P.: Assimilating remotely sensed snow observations into a macroscale hydrology model, Adv. Water Resour., 29(6), 872–886, doi:10.1016/j.advwatres.2005.08.004, 2006.

Arnold, J. G., Srinivasan, R., Muttiah, R. S. and Williams, J. R.: Large Area Hydrologic Modeling and Assessment Part I: Model Development, J. Am. Water Resour. Assoc., 34(1), 73–89, doi:10.1111/j.1752-1688.1998.tb05961.x, 1998.

Bergeron, J., Royer, A., Turcotte, R. and Roy, A.: Snow cover estimation using blended MODIS and AMSR-E data for
15 improved watershed-scale spring streamflow simulation in Quebec, Canada, Hydrol. Process., 28(16), 4626–4639, doi:10.1002/hyp.10123, 2014.

~~Brocca, L., Moramarco, T., Melone, F., Wagner, W., Hasenauer, S. and Hahn, S.: Assimilation of Surface- and Root-Zone ASCAT Soil Moisture Products Into Rainfall-Runoff Modeling, IEEE Trans. Geosci. Remote Sens., 50(7), 2542–2555, doi:10.1109/TGRS.2011.2177468, 2012.~~

20 Charbonneau, R., Fortin, J.-P. And Morin, G.: The CEQUEAU Model: Description and Examples of its use in Problems Related to Water Resource Management / Le modèle CEQUEAU: description et exemples d'utilisation dans le cadre de problèmes reliés à l'aménagement, Hydrol. Sci. Bull., 22(1), 193–202, doi:10.1080/02626667709491704, 1977.

Chen, H., Yang, D., Hong, Y., Gourley, J. J. and Zhang, Y.: Hydrological data assimilation with the Ensemble Square-Root-Filter: Use of streamflow observations to update model states for real-time flash flood forecasting, Adv. Water Resour., 59,
25 209-220, doi:10.1016/ j.advwatres.2013.06.010, 2013.

Clark, M. P., Slater, A. G., Barrett, A. P., Hay, L. E., McCabe, G. J., Rajagopalan, B. and Leavesley, G. H.: Assimilation of snow covered area information into hydrologic and land-surface models, Adv. Water Resour., 29(8), 1209–1221, doi:10.1016/j.advwatres.2005.10.001, 2006.

Day, G. N.: Extended Streamflow Forecasting Using NWSRFS, J. Water Resour. Plan. Manag., 111(2), 157–170, doi:10.1061/(ASCE)0733-9496(1985)111:2(157), 1985.

Dechant, C. M. and Moradkhani, H.: Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation, Hydrol. Earth Syst. Sci., 15(11), 3399–3410, doi:10.5194/hess-15-3399-2011, 2011.

Duan, Q. Y., Sorooshian, S. and Gupta, V.: Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models, Water Resour. Res., 28(4), 1015–1031, doi:10.1029/91wr02985, 1992.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., 99(C5), 10143, doi:10.1029/94JC00572, 1994.

Evensen, G.: The Ensemble Kalman Filter: Theoretical formulation and practical implementation, Ocean Dyn., 53(4), 343–367, doi:10.1007/s10236-003-0036-9, 2003.

Evensen, G. and van Leeuwen, P. J.: An Ensemble Kalman Smoother for Nonlinear Dynamics, Mon. Weather Rev., 128(6), 1852–1867, doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO;2, 2000.

Forman, B. a., Reichle, R. H. and Rodell, M.: Assimilation of terrestrial water storage from GRACE in a snow-dominated basin, Water Resour. Res., 48(1), 1–14, doi:10.1029/2011WR011239, 2012. Franz, K. J., Hogue, T. S., Barik, M. and He, M.: Assessment of SWE data assimilation for ensemble streamflow predictions, J. Hydrol., 519, 2737–2746, doi:10.1016/j.jhydrol.2014.07.008, 2014.

Franz, K. J., Hogue, T. S. and Sorooshian, S.: Snow Model Verification Using Ensemble Prediction and Operational Benchmarks, J. Hydrometeorol., 9(6), 1402–1415, doi:10.1175/2008JHM995.1, 2008.

Hall, D. K. and Riggs, G. A.: Accuracy assessment of the MODIS snow products, Hydrol. Process., 21(12), 1534–1547, doi:10.1002/hyp.6715, 2007.

Hall, D. K., Riggs, G. A., Salomonson, V. V, DiGirolamo, N. E. and Bayr, K. J.: MODIS snow-cover products, Remote Sens. Environ., 83(1-2), 181–194, doi:10.1016/S0034-4257(02)00095-0, 2002.

He, M., Hogue, T. S., Margulis, S. A. and Franz, K. J.: An integrated uncertainty and ensemble-based data assimilation approach for improved operational streamflow predictions, Hydrol. Earth Syst. Sci., 16(3), 815–831, doi:10.5194/hess-16-815-2012, 2012.

Houtekamer, P. L. and Mitchell, H. L.: Data assimilation using an ensemble Kalman filter technique, Mon. Weather Rev., 126(3), 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2, 1998.

Houtekamer, P. L. and Mitchell, H. L.: A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation, Mon. Weather Rev., 129(1), 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2, 2001.

Martinec, J.: Snowmelt - Runoff Model For Stream Flow Forecasts, Nord. Hydrol., 6(3), 145–154, 1974.

De Lannoy, G. J. M., Reichle, R. H., Arsenault, K. R., Houser, P. R., Kumar, S., Verhoest, N. E. C. and Pauwels, V. R. N.: Multiscale assimilation of Advanced Microwave Scanning Radiometer-EOS snow water equivalent and Moderate Resolution Imaging Spectroradiometer snow cover fraction observations in northern Colorado, Water Resour. Res., 48(1), 1–17, doi:10.1029/2011WR010588, 2012.

Lee, H., Seo, D. J. and Koren, V.: Assimilation of streamflow and in situ soil moisture data into operational distributed hydrologic models: Effects of uncertainties in the data and initial model soil moisture states, Adv. Water Resour., 34(12), 1597–1615, doi:10.1016/j.advwatres.2011.08.012, 2011.

Liu, Y. and Gupta, H. V.: Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, Water Resour. Res., 43(7), 1–18, doi:10.1029/2006WR005756, 2007.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., Seo, D. J., Schwanenberg, D., Smith, P., Van Dijk, a. I. J. M., Van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O. and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities, Hydrol. Earth Syst. Sci., 16(10), 3863–3887, doi:10.5194/hess-16-3863-2012, 2012.

Marcil, G.-K., Leconte, R., Trudel, M.: Using Remotely Sensed MODIS Snow Product for the Management of Reservoirs in a Mountainous Canadian Watershed, Water Resour. Manag., Accepted, 2016.

McMillan, H. K., Hreinsson, E. O., Clark, M. P., Singh, S. K., Zammit, C. and Uddstrom, M. J.: Operational hydrological data assimilation with the recursive ensemble Kalman filter, Hydrol. Earth Syst. Sci., 17(1), 21–38, doi:10.5194/hess-17-21-2013, 2013.

Nagler, T., Rott, H., Malcher, P. And Muller, F.: Assimilation of meteorological and remote sensing data for snowmelt runoff forecasting, Remote Sens. Environ., 112(4), 1408–1420, doi:10.1016/j.rse.2007.07.006, 2008.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, J. Hydrol., 10(3), 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

5   De Roo, A. P. J., Gouweleeuw, B., Thielen, J., Bartholmes, J., Bongioannini-Cerlini, P., Todini, E., Bates, P. D., Horritt, M., Hunter, N., Beven, K., Pappenberger, F., Heise, E., Rivin, G., Hils, M., Hollingsworth, A., Holst, B., Kwadijk, J., Reggiani, P., Van Dijk, M., Sattler, K. and Sprokkereef, E.: Development of a European flood forecasting system, Int. J. River Basin Manag., 1(1), 49–59, doi:10.1080/15715124.2003.9635192, 2003.

Roy, A., Royer, A. and Turcotte, R.: Improvement of springtime streamflow simulations in a boreal environment by
10   incorporating snow-covered area derived from remote sensing data, J. Hydrol., 390(1-2), 35–44, doi:10.1016/j.jhydrol.2010.06.027, 2010.

Samuel, J., Coulibaly, P., Dumedah, G. and Moradkhani, H.: Assessing model state and forecasts variation in hydrologic data assimilation, J. Hydrol., 513, 127–141, doi:10.1016/j.jhydrol.2014.03.048, 2014.

Slater, A. G. and Clark, M. P.: Snow Data Assimilation via an Ensemble Kalman Filter, J. Hydrometeorol., 7(3), 478–493,
15   doi:10.1175/JHM505.1, 2006.

Tang, Q. and Lettenmaier, D. P.: Use of satellite snow-cover data for streamflow prediction in the Feather River Basin, California, Int. J. Remote Sens., 31(14), 3745–3762, doi:10.1080/01431161.2010.483493, 2010.

Trudel, M., Leconte, R. and Paniconi, C.: Analysis of the hydrological response of a distributed physically-based model using post-assimilation (EnKF) diagnostics of streamflow and in situ soil moisture observations, J. Hydrol., 514, 192–201,
20   doi:10.1016/j.jhydrol.2014.03.072, 2014.

U. S. Army Corps of Engineers: Snow Hydrology, Summary report of the snow investigation, North Pacific Division, Portland, Oregon, 1956.

Weerts, A. H. and El Serafy, G. Y. H.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resour. Res., 42(9), 1–17, doi:10.1029/2005WR004093, 2006.

25   Xie, X. and Zhang, D.: Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter, Adv. Water Resour., 33(6), 678–690, doi:10.1016/j.advwatres.2010.03.012, 2010.

Table 1: Overview of multivariate DA scenarios.

Method	Streamflow DA updates:		SWE DA updates:	
	VOL + GWL?	SWE?	VOL + GWL?	SWE?
VG-S	yes	no	no	yes
VG2-S	yes	yes	no	yes
VG-S2	yes	no	yes	yes
VG2-S2	yes	yes	yes	yes

5
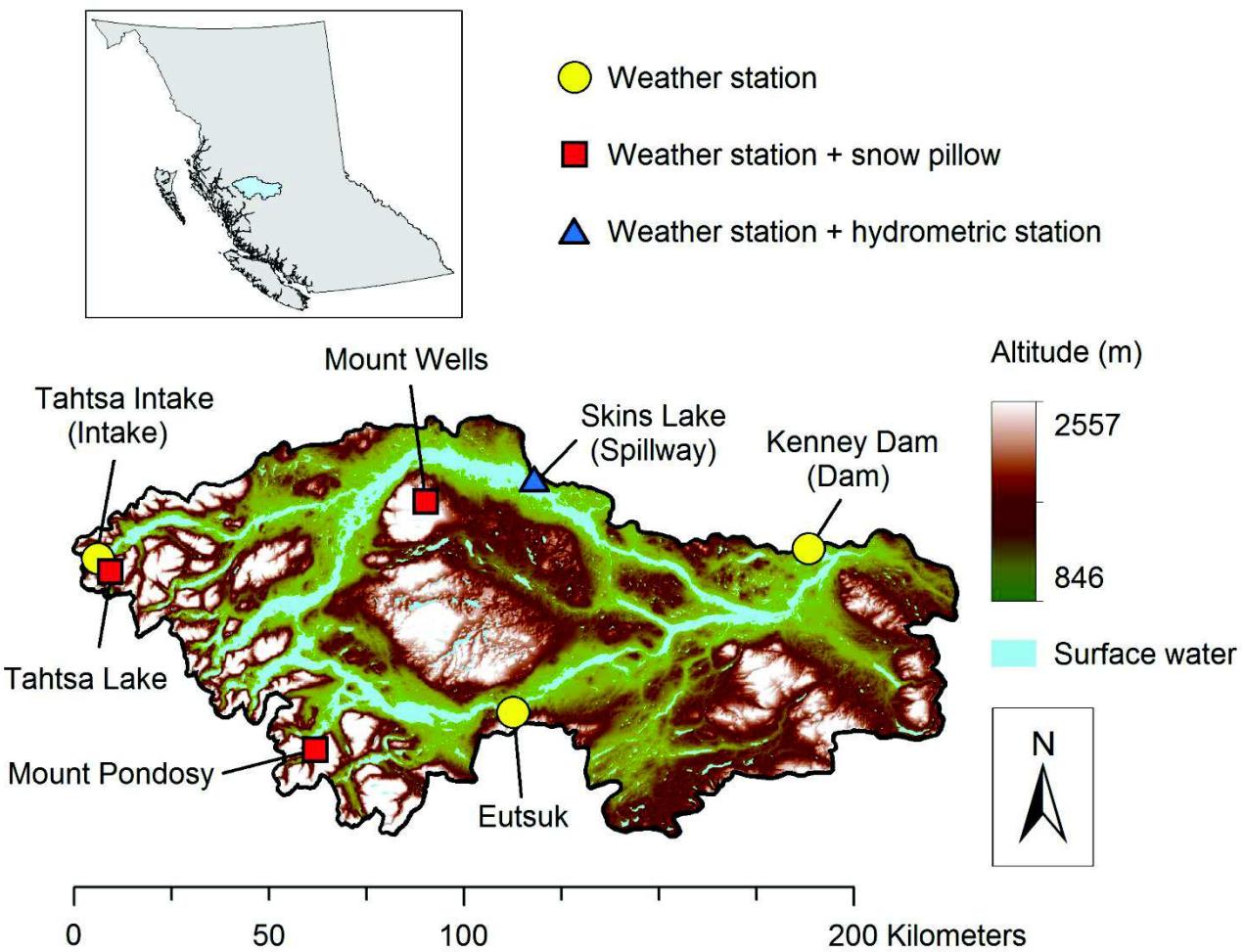
Figure 1: The Nechako watershed and the locations of weather stations, snow pillows and a hydrometric station. All of these contain at least daily weather data. The outlet is considered to be at the spillway, located at the blue triangle. The intake is located at the Tahtsa Intake weather station (westernmost yellow circle).
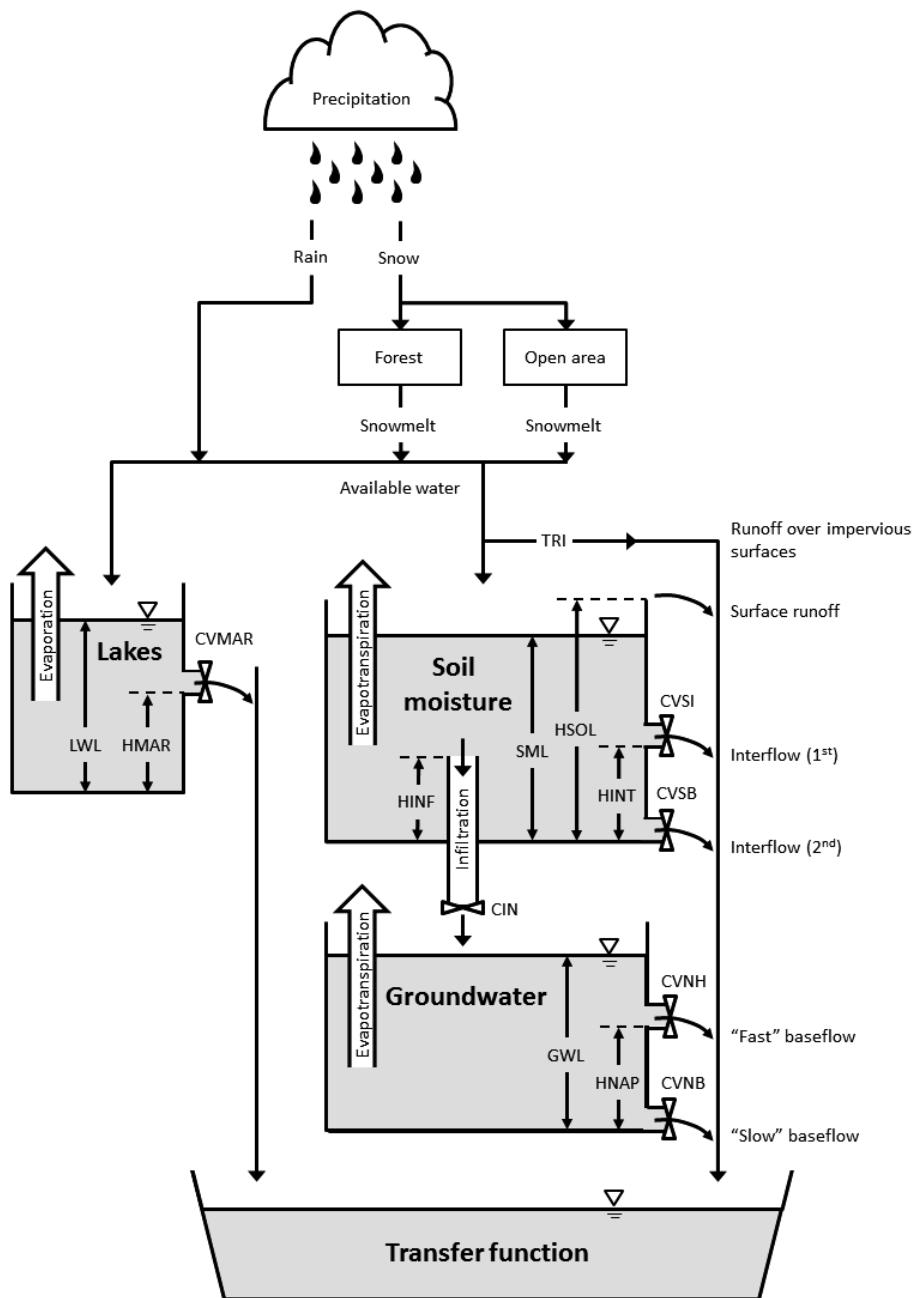
Figure 2: Diagram of the processes included in CEQUEAU's production function.
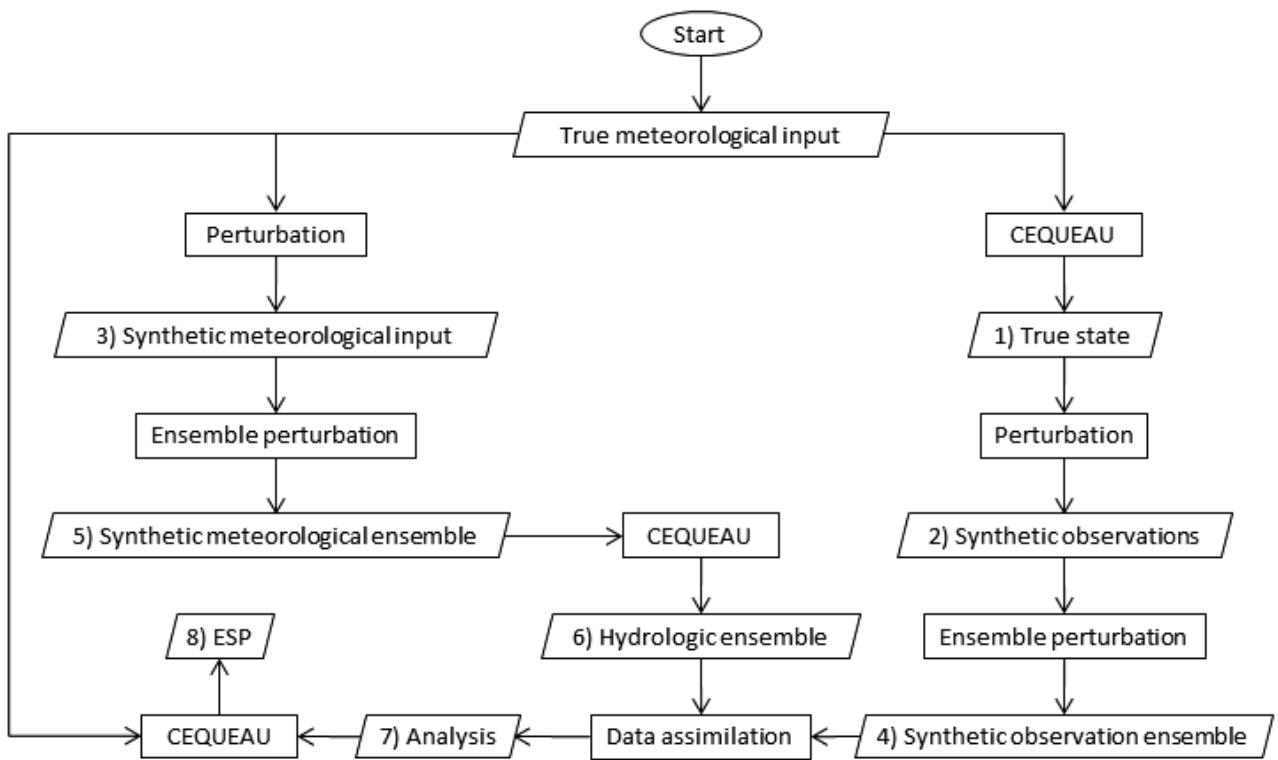
Figure 3: Flowchart for the production of ensemble streamflow predictions obtained from various data assimilation scenarios.
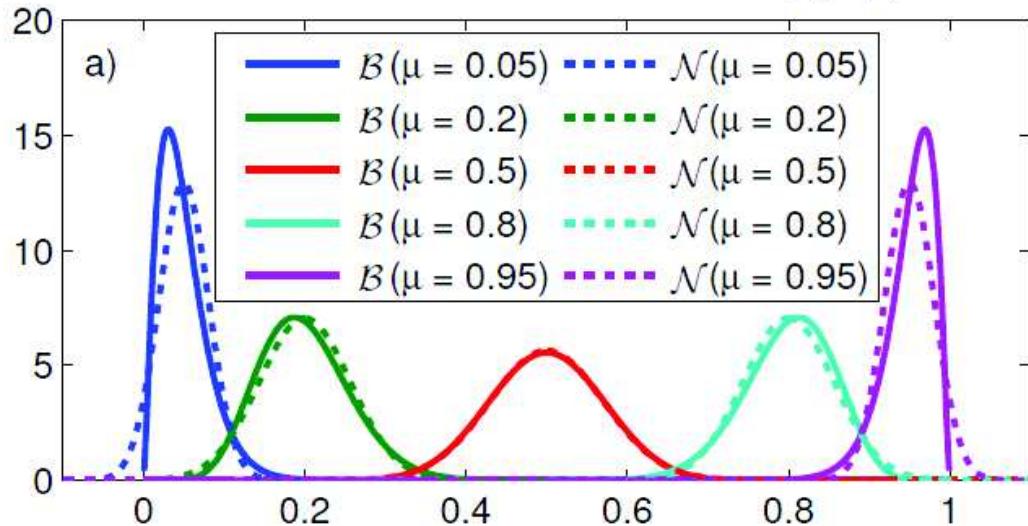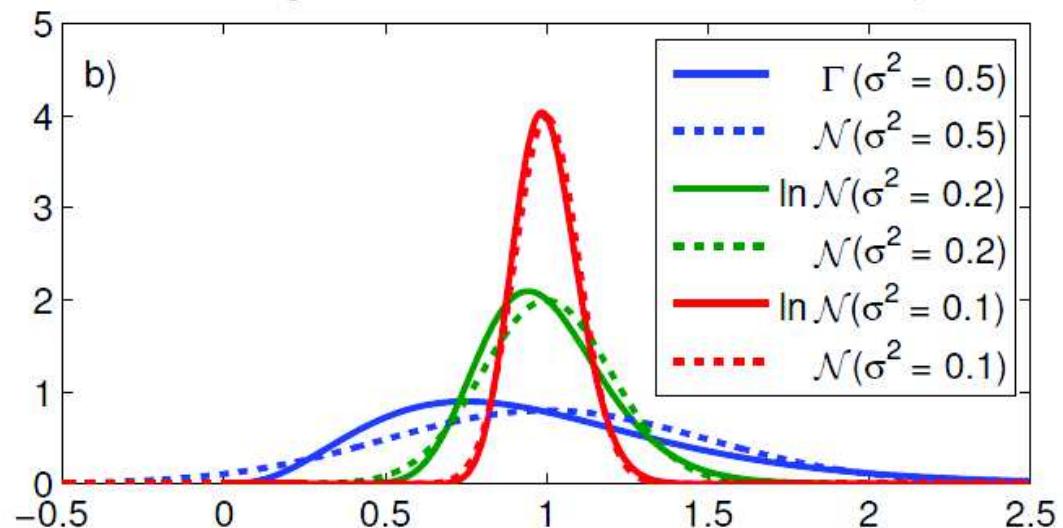
Figure 4. Examples of a) beta and b) gamma and lognormal distribution compared with their analogous normal distributions using the same mean (μ) and variance (σ²).
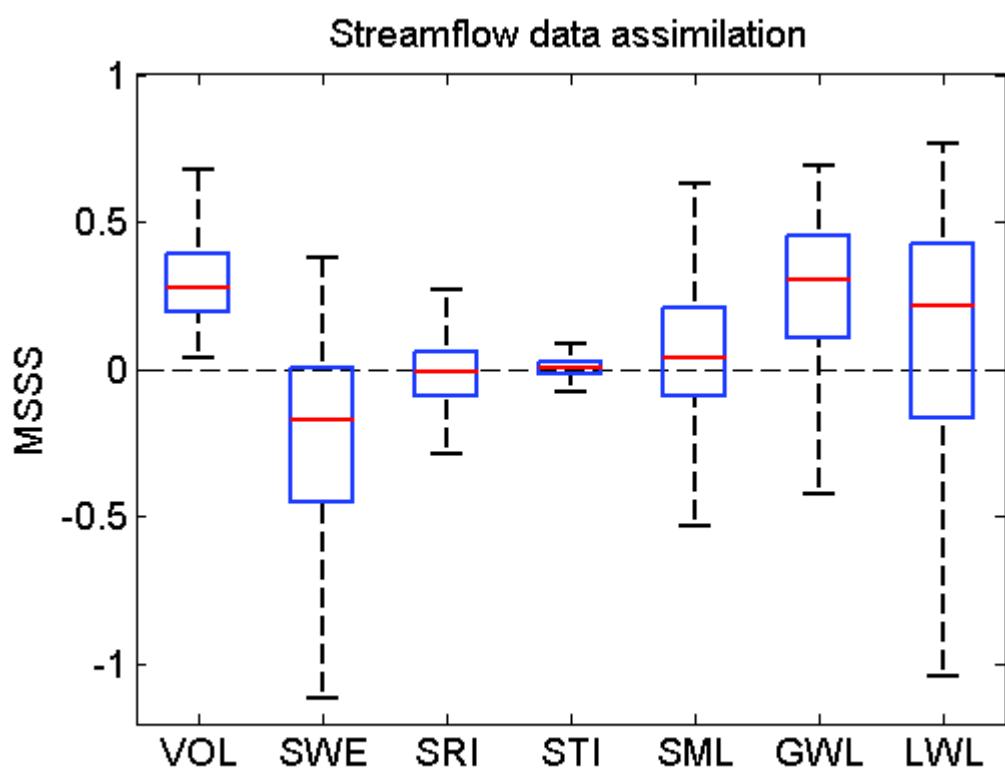
Figure 54: Box plot of the Mean Square Skill Score for each variable when assimilating streamflow at the outlet. The open loop is used as a reference. Outliers are not shown for visibility purposes.
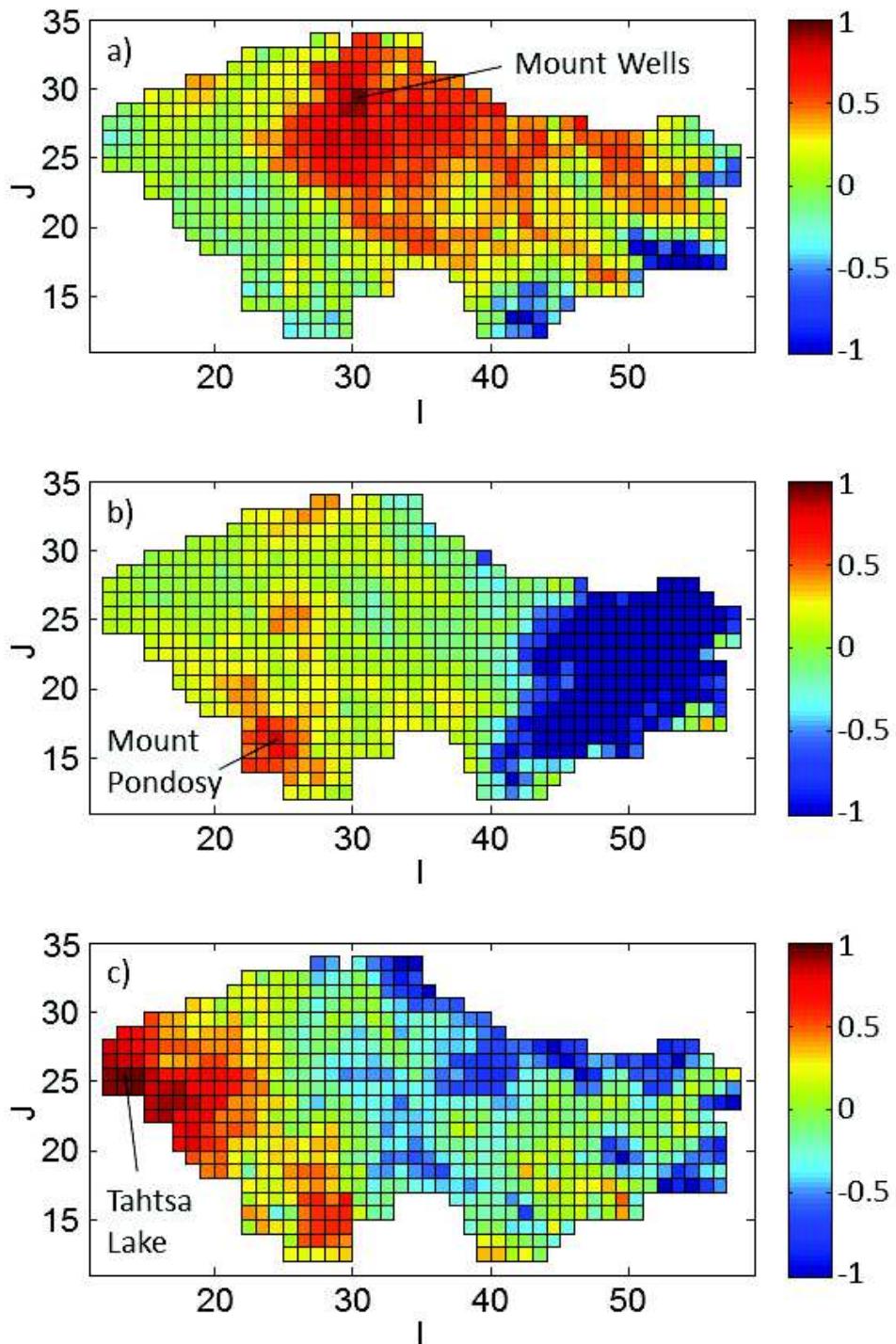
Figure 65: Distribution of the Mean Square Skill Score of SWE over the watershed when assimilating SWE located at a) Mount Wells, b) Mount Pondosy and c) Tahtsa Lake. The open loop is used as a reference. Values below -1 are cut off from the legend.
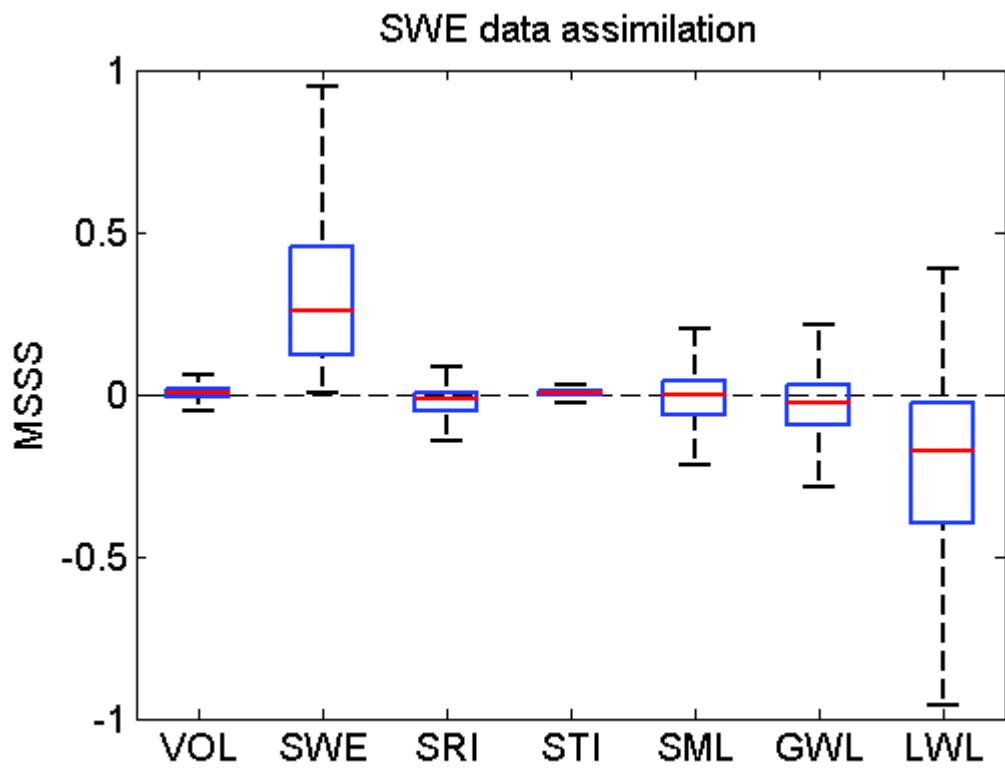
Figure 76: Box plot of the Mean Square Skill Score for each variable when assimilating SWE from all three snow pillow locations. The open loop is used as a reference. Outliers are not shown for visibility purposes.
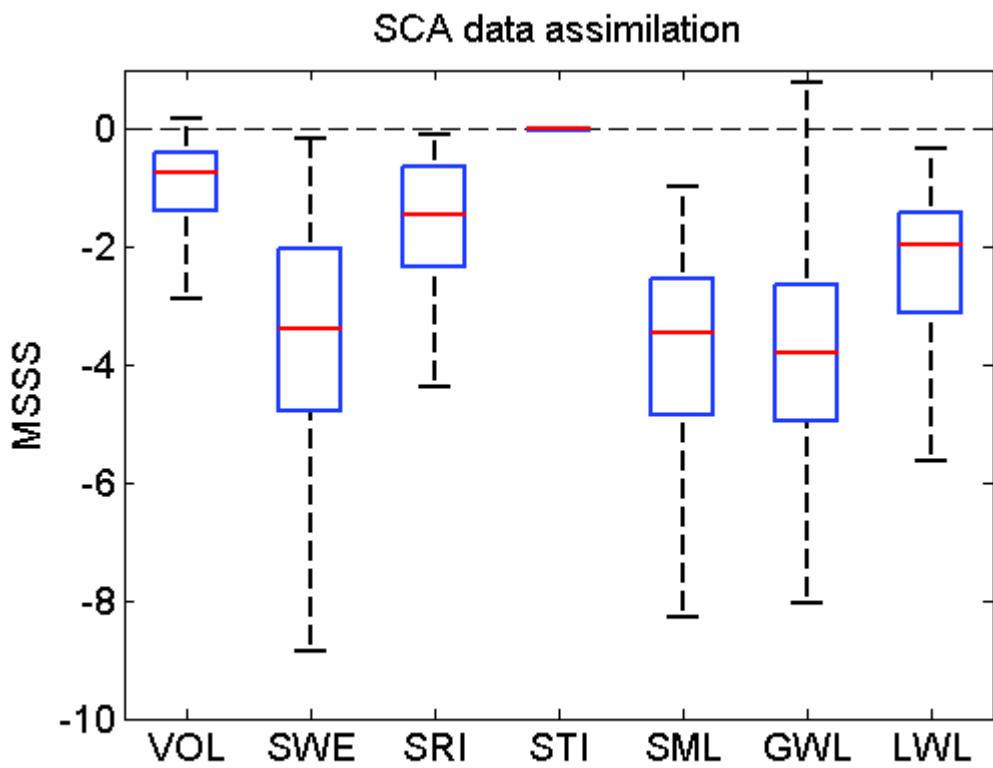
Figure 87: Box plot of the Mean Square Skill Score for each variable when assimilating basin-wide snow cover area. The open loop is used as a reference. Outliers are not shown for visibility purposes.
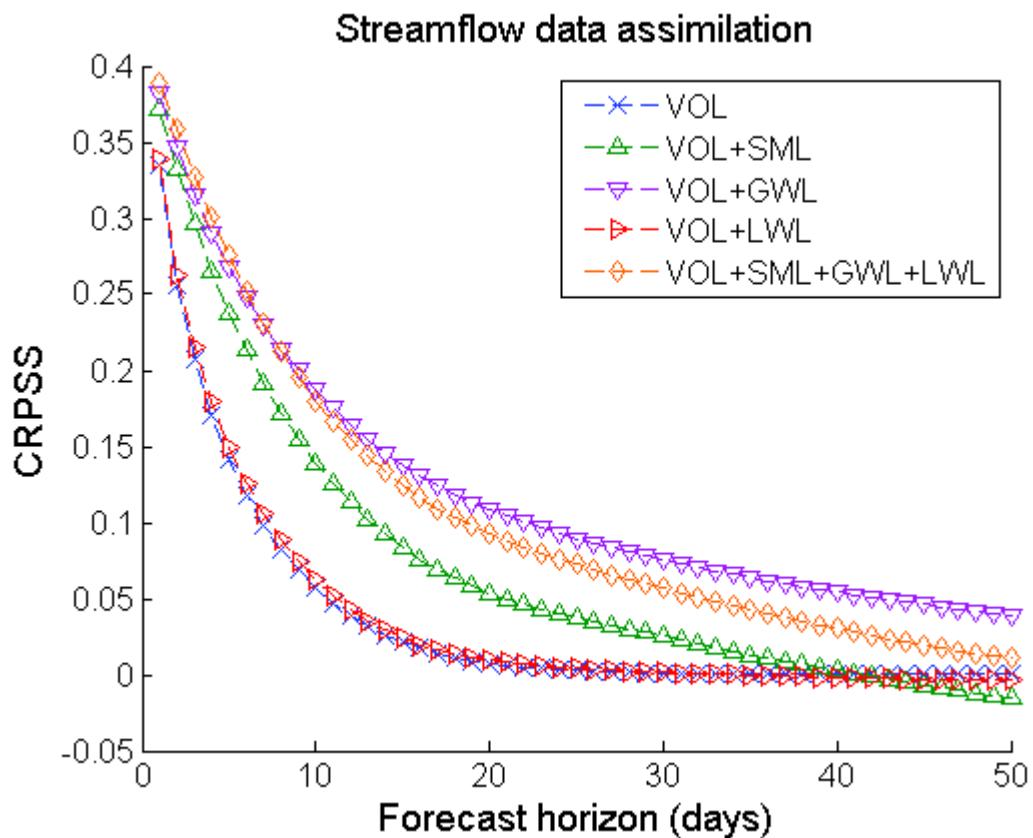
Figure 98: Continuous Rank Probability Skill Score of the streamflow ensemble when assimilating streamflow at the outlet. The open loop is used as a reference. The forecast horizon varies from 1 to 50 days.
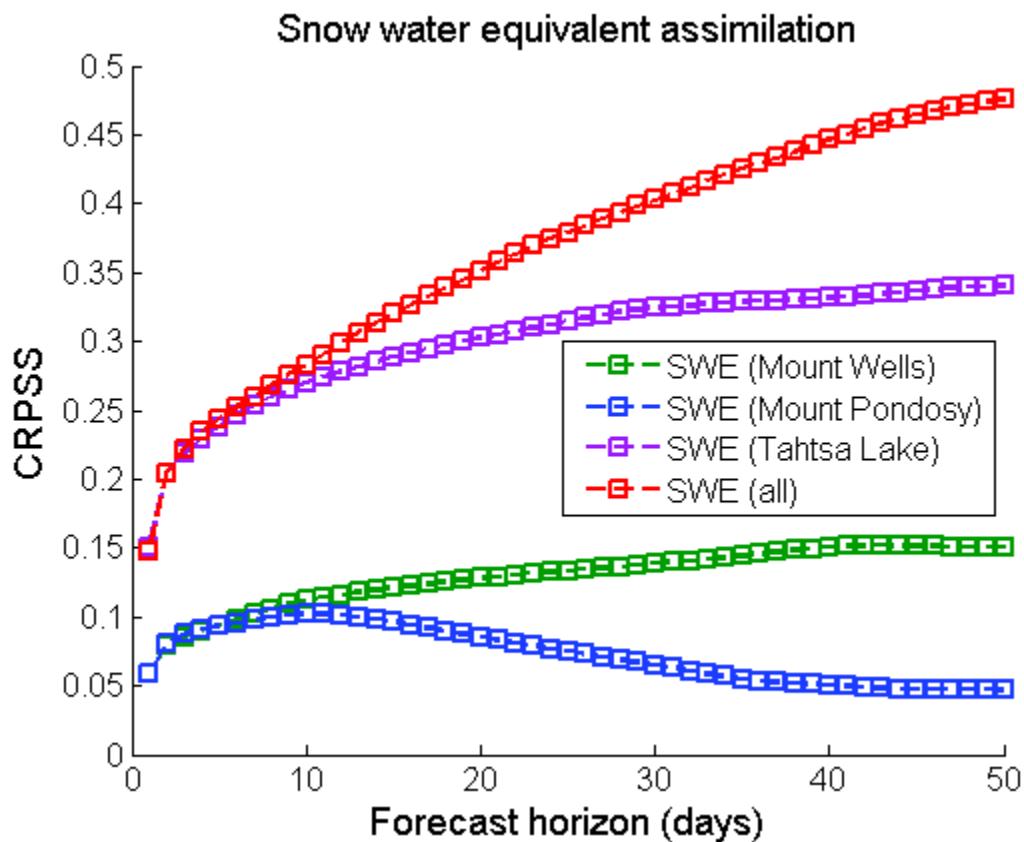
Figure 109: Continuous Rank Probability Skill Score of the streamflow ensemble when assimilating SWE from all three snow pillow locations. The open loop is used as a reference. The forecast horizon varies from 1 to 50 days.
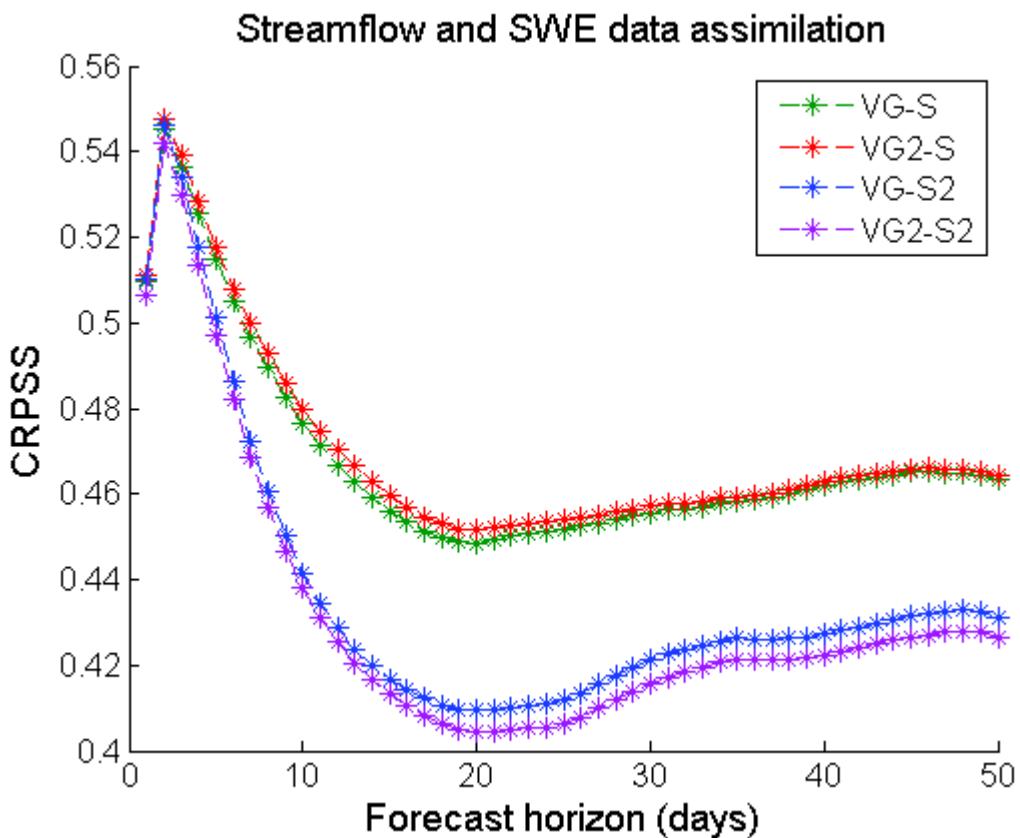
Figure 11~~10~~: Continuous Rank Probability Skill Score of the streamflow ensemble when assimilating streamflow at the outlet and SWE from all three snow pillow locations. The open loop is used as a reference. The forecast horizon varies from 1 to 50 days. Lack of parentheses indicates that the variable is affected by both types of observations.
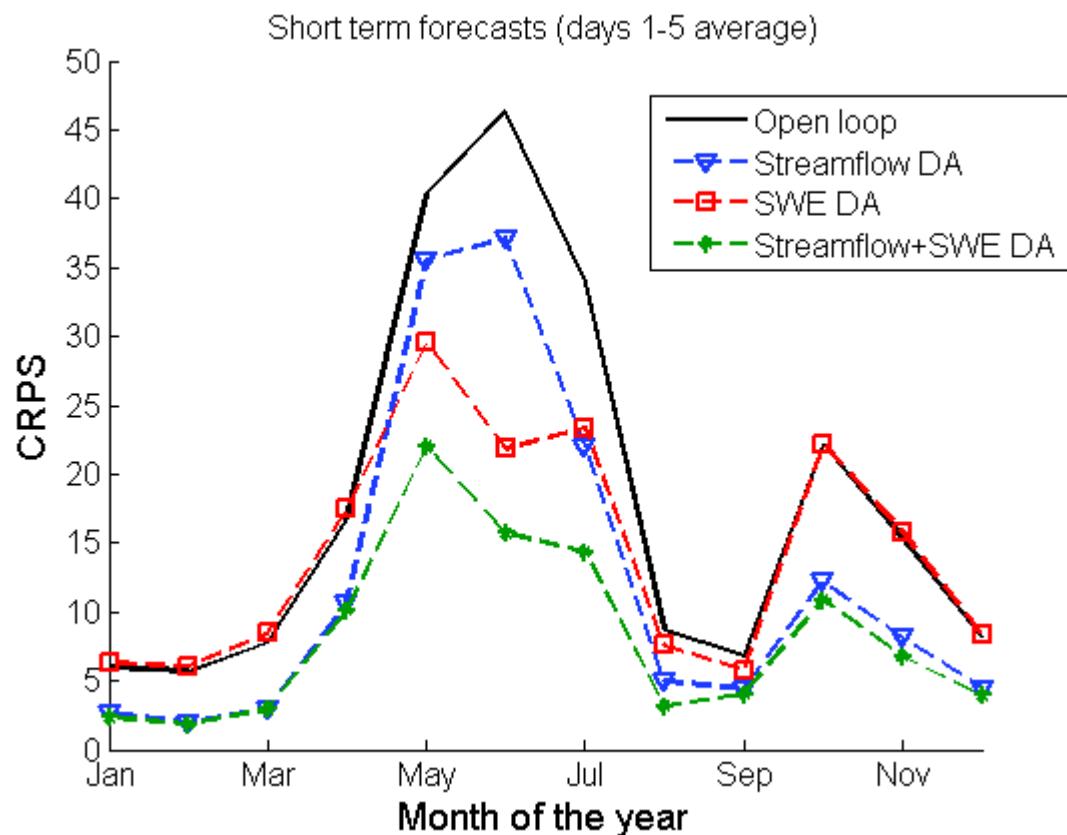
Figure 12₁₁: Continuous Rank Probability Score for short-term forecasts (average of forecast horizons 1 through 5 days) of various data assimilation scenarios as a function of the month of the year.
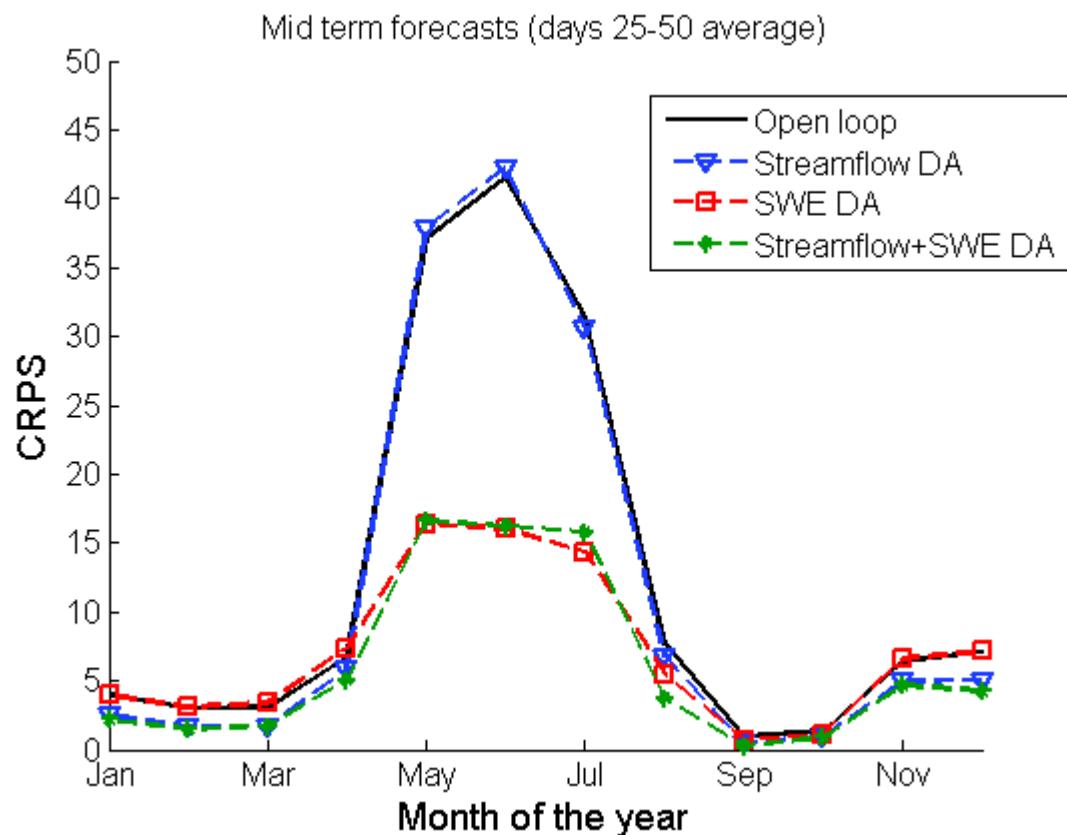
Figure 13~~12~~: Continuous Rank Probability Score for mid-term forecasts (average of forecast horizons 25 through 50 days) of various data assimilation scenarios as a function of the month of the year.

5