Particle Filter data assimilation of streamflow in basins with seasonal snow for initializing short- to medium-range streamflow forecasts

Elizabeth A. Clark

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Bart Nijssen, Chair

Andrew Wood

Faisal Hossain

Program Authorized to Offer Degree:

Civil and Environmental Engineering

University of Washington

Abstract

Particle Filter data assimilation of streamflow in basins with seasonal snow for initializing short- to medium-range streamflow forecasts

Elizabeth A. Clark

Chair of the Supervisory Committee:
Associate Professor WOT Bart Nijssen
Civil and Environmental Engineering

Although short- to medium-range streamflow forecasting is vital for many water management decisions the quality of streamflow forecasts in the United States has not improved over time. A possible explanation for this is that the current forecast systems, which rely on manual intervention on the part of forecasters, do not easily allow for testing of potential system upgrades and new methodologies. To transition from a semi-manual procedure to a fully automated procedure—and thereby allow hindcasting experiments to test new methods—forecast system must use an automated data assimilation (DA) framework. In this dissertation, I evaluate the capabilities of one DA method—the particle filter (PF) —for the assimilation of streamflow observations in basins with seasonal snow cover. Very few studies have explored the use of DA

based solely on streamflow to update snow states. Studying DA based solely on streamflow in basins with seasonal snow cover is important because streamflow observations are much more widely available than observations of snow cover extent, depth or water content. I first use a synthetic experiment to examine the impacts of such DA on streamflow, snow states, and soil moisture states. PF-DA almost always improves simulated soil moisture and streamflow in two Pacific Northwest basins with seasonal snow, but it degrades the quality of snow water equivalent estimates during the mid-winter and in the basin for which snow is less of a control on runoff. Next, I evaluate to what extent the improved initial hydrologic conditions lead to improvements in the 1- to 7-day lead-time forecasts. I find that PF of streamflow observations in basins with seasonal snow improves forecast performance in terms of accuracy during the spring and summer and in terms of reliability during spring and fall. Finally, I propose an Analog Resampling (AR) method, for use in the PF, that shows potential for expanding the spread of particles when the particle sample degenerates (i.e., when most weight is assigned to few particles). A set of exploratory analyses using AR suggests areas for future development of this method.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

My sincerest gratitude goes to my academic advisor Bart Nijssen who took a chance on me and provided the guidance and support that I needed to complete this dissertation. I would also like to thank the members of my doctoral exam committee, including Andrew Wood, Faisal Hossain, Erkan Istanbullouglu, my graduate school representative Jeffrey Richey, and my general exam graduate school representative Luanne Thompson. The over-the-loop demonstration project was led by Andrew Wood (PI), Bart Nijssen (PI), and Martyn Clark (co-PI), and it was much improved by feedback and support from Levi Brekke, Ken Nowak, and Jeff Arnold. I was lucky to work with several supportive, hard-working, and creative members of the NCAR Research Applications Laboratory, including Pablo Mendoza, Manabendra Saharia, Ethan Gutmann, Joseph Hamman, Andrew Newman, and Martyn Clark. I would like to thank the many members of the UW Computational Hydrology Research Group, led by Bart Nijssen, and of the former UW Land Surface Hydrology Research Group, led by Dennis Lettenmaier, especially Marketa McGuire and Jenny Adam, who started me off with my first hydrologic modeling experiments. I would also like to give a special thanks to Konstantinos Andreadis and Nathalie Voisin who have been mentors, colleagues, and friends throughout this endeavor. Finally, I am grateful to my family for continuing to say, "I'm proud of you," throughout the process and to my emotional support team, especially Emma Levitt, Julia Pollak, Elliot Grunewald, and Elizabeth Cassel.

This dissertation is partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement NA10OAR4320148 (2010-2015) and NA15OAR4320063 (2015-2020), Contribution No. 2017-093. Funding was also provided by the U.S. Bureau of Reclamation under Cooperative Agreement R11AC80816, and by the U.S. Army

Corps of Engineers Climate Preparedness and Resilience Program (Award Number 1254557). This work was facilitated though the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington. GEFS reforecasts were processed at Computational and Information Systems Laboratory (2012). Precipitation and temperature observations used for model spin-up can be obtained from the Global Historical Climatology Network at http://doi.org/10.7289/V5D21VHZ. Daily no-regulation, no-irrigation (NRNI) flows were used for model calibration and are available from the Bonneville Power Administration at http://www.bpa.gov/power/streamflow/default.aspx.

DEDICATION

For my dad, Ronald A. Clark.

Chapter 1. INTRODUCTION

1.1 BACKGROUND

Short- to medium-range streamflow forecasts (at lead times of one to seven days) provide invaluable information for the management of water resources for a wide variety of applications, including flood control, irrigation, municipal and industrial use, environmental protection, and recreation. In the United States, official streamflow forecasts are issued by the National Weather Service (NWS) River Forecasting Centers (RFCs). These forecasts incorporate many sources of information, including hydrologic model simulations and numerical weather predictions. Model simulations are inherently erroneous due to model structural errors, parameter errors, and errors in forcing data. To minimize these errors, methods have been developed to incorporate additional information as it becomes available. These methods are known as data assimilation, and they aim to intelligently combine model simulations and observations, accounting for errors in each.

Forecasters perform a direct data assimilation process in which they manually adjust model-based forecasts to account for discrepancies with observations [*McEnery et al.,* 2005]. The primary benefit of this approach is that it directly leverages operator experience. One disadvantage to this approach is that it prohibits the use of past or current forecast performance as a predictor of future forecast performance and as a result, does not allow for objective evaluation of competing forecasting strategies. The reason for this is that personnel, data systems, and operator experience all change over time, and these factors impact the quality of forecasts that rely on manual intervention. The manual adjustments also make it impossible to evaluate changes to the system because they eliminate the possibility of hindcasting experiments [*Raff et al.,* 2013]. The idea behind a hindcasting experiment is to perform a series of "forecasts"

for a historical period using the same techniques used in a real-time forecasting system. These hindcasts can then be used to develop metrics regarding forecast performance and to develop statistical post-processing techniques that correct for systematic errors in the forecast. *Welles et al.* [2007] argue that systematic verification of numerical weather predictions has directed research towards approaches that objectively improve forecasts. They suggest that the hydrologic forecasting community could benefit from a similar system of verification.

Automation of the forecasting process, including data assimilation, would allow forecasters to create a long record of past forecasts, over a range of climatic conditions and hydrologic events, that are consistent with the present operational system. Although an automated forecasting workflow, by definition, requires the automation of data, that step has yet to be adapted operationally in the United States for several reasons. *Hartman* [2014] explains four of these reasons as follows: 1) the changes made to improve short term forecasts may worsen long-term forecasts, and vice versa; 2) improved performance of a hydrologic model does not necessarily lead to improved forecasts; 3) forecasters may see data assimilation as a "black box" that they do not trust; and 4) forecasting centers may be concerned that future forecasters will no longer understand how the models operate, which may impact their ability to interpret the forecasts. To address the first three of these concerns, we developed a test-bed automated streamflow forecasting processes known as the System for Hydromet Analysis, Research and Prediction (SHARP) [*Wood et al*., 2016; *Mendoza et al.*, 2017]. The approach demonstrated by SHARP is also known as 'over-the-loop' forecasting in that the forecaster interprets the forecast once it is issued and interacts with the forecasting system through methods development and systematic testing. This dissertation focuses on a single aspect of SHARP—data assimilation—however, the demonstration project also involves automation of

meteorological forcing generation, downscaling of meteorological forecasts, streamflow forecast post-processing, and seasonal forecasts.

Among the approaches to automated data assimilation, particle filters (PFs) [*Kitagawa et al.,* 1996] are popular for hydrologic applications because they work in nonlinear systems [*Arulampalam et al*., 2002] and maintain internally consistent model states and parameters [*Salamon and Feyen*, 2010]. While PFs based solely on the assimilation of streamflow observations have been applied in several hydrologic studies [*Dumedah and Coulibaly*, 2013; *Moradkhani et al*., 2005, 2006; *Weerts and El Serafy*, 2006], they have not been used to update snow states in basins that have seasonal snow. Chapters 2 and 3 of this dissertation address the implications of applying a PF based solely on streamflow observations in two basins with seasonal snow. PFs have two well-known problems: 1) sample degeneracy and 2) its consequence, sample impoverishment [*Li et al.,* 2014]. Chapters 2 and 3 show this to be a problem in basins with seasonal snow, particularly when the model ensemble is biased relative to observed streamflow. Existing solutions to correct for sample impoverishment [*Li et al.,* 2014] do not preserve the internal consistency of states (i.e., they have the potential to produce soil moisture and snow water equivalent time series that could not reasonably co-exist under current conditions). Chapter 4 proposes a new methodology based on historical analogs to limit sample impoverishment while preserving the internal consistency of states.

1.2 Objectives and Research Questions

This dissertation explores the use of PF data assimilation to improve the simulated hydrologic states used to initialize streamflow forecasts. I focus on the application of the Sequential Importance Resampling Particle Filter (SIR-PF) [*Arumpalam et al*., 2002; *Doucet et al.,* 2001] to assimilate streamflow observations to update simulated soil moisture, snow water equivalent,

and streamflow. The primary objective of this dissertation is to improve our understanding of the implications of and potential for PF data assimilation based solely on streamflow observations in basins with seasonal snow cover.

Research Questions

1) What effect does PF data assimilation based solely on streamflow observations have on simulated distributions of streamflow, soil moisture, and snow water equivalent?
2) To what degree does SIR-PF data assimilation based solely on streamflow improve short- to medium-range (1-7 day) forecasts, in terms of accuracy and reliability?
3) Can the use of hydrologic analogs during periods of poor ensemble performance improve PF data assimilation based solely on streamflow?

1.3 APPROACH

I address each of these three research questions in Chapters 2, 3, and 4, respectively. For Chapters 2 and 3, I compare results in two study basins in the Pacific Northwest, United States. The first of these is larger and primarily receives its precipitation in the form of snow, while the second, smaller basin receives a mixture of rain and snow. In Chapter 2, I perform a set of PF experiments that use synthetic observations of streamflow taken from a model simulation that is treated as "true". This allows me to compare the simulated distribution of streamflow, soil moisture and snow water equivalent from the SIR-PF with the "true" states (Question 1). I use the Continuous Rank Probability Score (CRPS) [*Hersbach*, 2000] to evaluate each component probabilistically. I compare the SIR-PF results with two baselines: 1) a simulation that uses no data assimilation and 2) simulations that use a simpler ad-hoc non-sequential PF. In Chapter 3, I use real observations of naturalized streamflow in a SIR-PF to simulate hydrologic states for seven water years. I then use these states to initialize 1- to 7-day streamflow forecasts. I compare

the resulting forecasts—in terms of CRPS, α reliability [*Renard et al*., 2010], mean absolute error, percent bias, and correlation—with forecasts that were initialized from simulations that did not use data assimilation (Question 2). Finally, I address Question 3 in Chapter 4, which describes two proposed approaches to resampling particles from historical analogs and applies them to five types of hydroclimatic conditions. The simulated hydrologic states and streamflow resulting from that exploratory analysis are compared to observed streamflow, states and streamflow simulated without data assimilation, and states and streamflow simulated using the SIR approach to resampling.

Chapter 2. ASSESSING ENSEMBLE PARTICLE FILTERS FOR THE ESTIMATION OF MODEL STATES FOR STREAMFLOW FORECASTING

This chapter is in revision for submission to *Water Resources Research* as

Clark, E. A., A. W. Wood, and B. Nijssen. Assessing ensemble particle filters for the estimation of model states for streamflow forecasting. *Water Resources Research*, in revision.

Abstract

Streamflow forecasts produced by the National Weather Service River Forecast Centers incorporate ancillary observations through supervised manual data assimilation (DA) to adjust model states used to initialize streamflow forecasts. Automation of DA is regarded as a critical step for advancement of operational streamflow prediction, as it improves consistency and reproducibility. The particle filter (PF) is an attractive method of DA because it does not require adjustments to model states and can be used for non-linear and non-Gaussian systems. Using a synthetic experiment, in which one simulation of states and streamflow is regarded as truth, we assess the ability of a recursive (sequential) PF versus a simpler non-sequential PF to maintain skillful moisture state ensembles while assimilating only streamflow. The resultant ensembles of daily hydrologic states from water year 2006 to 2015 in two watersheds – the Green River Basin upstream of Howard Hanson Dam and the South Fork Flathead River upstream of Hungry Horse Dam – were evaluated in terms of Continuous Rank Probability Score (CRPS). The sequential PF produced the best streamflow distributions at both sites; however, the difference in CRPS between sequential and non-sequential filters decreased when observations were given higher

weight relative to the model. Soil moisture distributions improved for all filters in both basins, due to the tight coupling of soil moisture to streamflow, but DA only consistently improved the snow water equivalent distribution in the Hungry Horse basin, underscoring the relatively stronger influence of snow melt on flow generation in that watershed versus the Howard Hanson basin.

2.1 INTRODUCTION

Short- to medium-range streamflow forecasts inform daily water system operational decisions throughout the world. These forecasts are uncertain due to a combination of unknown initial hydrologic conditions (IHCs), measurement errors, data interpolation errors, hydrologic model parameter uncertainty and structural errors, and uncertain quantitative precipitation forecasts and downscaling techniques [*Cloke and Pappenberger*, 2009; *Wood and Lettenmaier*, 2008]. In this study, we produce an ensemble of simulated watershed moisture states that reflect the uncertainties in initial conditions and that can be used to initialize model streamflow forecasts. One approach to improve the estimated distribution of IHCs is to incorporate observations into simulations through data assimilation (DA). The current operational streamflow forecasts from the National Weather Service (NWS) River Forecast Centers (RFCs), which are distributed via the Advanced Hydrologic Prediction Service (AHPS), uses a predominantly manual DA approach for hydrologic state updating [*McEnery et al.*, 2005]. Because manual DA leverages the forecasters' experience and judgment, the decision-making process evolves over time due to changes in experience, data systems, and personnel. For this reason, manual DA undermines the consistency and reproducibility of forecasts, which in turn limits development opportunities that require these characteristics. For instance, reproducibility enables consistent hindcasts to support forecast verification, statistical post-processing, and the benchmarking of new forecasting

strategies against existing ones. Automated DA methods also enable the use of higher dimensional, more complex models as well as greater forecast frequency and more forecast locations. Three methods of automated DA have gained popularity in hydrologic studies: 1) the Kalman filter and its variants, 2) particle filters (PFs), and 3) variational methods [*McLaughlin*, 2002]. PFs are attractive to hydrologic modeling applications because they can be applied to nonlinear and non-Gaussian problems [*Arulampalam et al*., 2002] and because they do not require direct manipulation of model states [*Salamon and Feyen*, 2010].

Particle filters represent the posterior distribution of model states as an ensemble of individually weighted simulation outcomes ('particles'). As with other ensemble-based methods, each particle represents a plausible model state (soil moisture, snow water equivalent, temperature, etc.) and tracks its evolution as propagated in time by a model. Each particle has a weight based on its likelihood of representing the true state given respective uncertainties in the hydrologic model and observations. Many previous studies in hydrology [for example, *Moradkhani et al*., 2005, 2006; *Weerts and El Serafy*, 2006; *DeChant and Moradkhani*, 2011a, 2014; *Salamon and Feyen*, 2010] have applied PF using the method of Sequential Importance Resampling (SIR) to improve simulation performance. SIR, which is derived directly from Bayes theorem, combines a method called Sequential Importance Sampling (SIS), which estimates the posterior distribution based on the likelihood of a given state, with a resampling step to avoid sample degeneracy (see section 2.2.2). Sample degeneracy occurs when only a few particles are given the majority of the weight, such that the particle distribution cannot represent the posterior distribution. SIS assumes that the true distribution of states is proportional to some known (importance) distribution [*Doucet et al.,* 2001]. In practice, it is common to use a pre-specified prior distribution as the importance distribution. The weights in SIS are calculated recursively,

which allows the prior likelihood of states to influence the posterior weights. Particle resampling removes particles that perform poorly and spawns additional particles from the states of the best performing particles. A downside of SIR is that it can be challenging to interpret the recursive weights to analyze why a certain particle is excluded from the filter ensemble on a given day.

In this paper, we focus on the practical implementation of the particle filter to estimate model states to improve, potentially, their application as IHCs for streamflow forecasting applications. We apply the SIR approach, and we also benchmark it through comparison to a non-sequential, non-recursive alternative to SIR that limits particle evaluation and weighting solely to the prior week's performance. At each assimilation time, the non-sequential particle filter evaluates and selects particles to retain, cull or resample, as with the SIR filter. There are four primary benefits to this approach: 1) intuitive weight interpretation; 2) greater user control; 3) less sample degeneracy; 4) potentially lower computational burden.

The objectives of this paper are two-fold. First, we aim to evaluate the marginal benefits of SIR versus limited-memory non-sequential particle filter implementations to estimate ensemble distributions of hydrologic states (snow water equivalent and soil moisture) and streamflow, based solely on the assimilation of streamflow observations. Second, we wish to understand more broadly the ability of a streamflow-only assimilation to improve other key model moisture states, since these are critical for initializing streamflow forecasts. The methods examined here are designed to be suitable for deployment into a fully automated streamflow forecast system.

The following section (2.2) describes the methodological approach. The data sources, study sites, and hydrologic model used in this study are described in section 2.3. Section 2.4 presents results for each of the experiments in terms of streamflow, soil moisture, and snow

water equivalent (SWE). A discussion of these results and their broader implications is included in section 2.5, and conclusions with an eye toward implementation are presented in section 2.6.

2.2 APPROACH

This study assimilates streamflow into an ensemble of watershed simulations. Because the focus of this paper is on the particle filter performance and its impact on watershed states, which are rarely directly measurable, we perform an experiment in which one member of the ensemble of model simulations is randomly selected as truth – i.e., providing synthetic 'observations' of streamflow, and hydrologic states (snow water equivalent and soil moisture). This experiment sheds light on the ability of DA to draw the model states closer to the model 'truth' while avoiding degeneracy.

The key elements of the approach and workflow used to simulate model states are described in section 2.2.1. A description of the elements common to all particle filter methods employed in this study is given in section 2.2.2. Section 2.2.3 presents the details of each of six experiments conducted, and section 2.2.4 describes the metrics used to assess the results of each experiment.

2.2.1 *Hydrologic model state ensemble*

In this study, we represent errors in simulated watershed moisture states and streamflow using an ensemble of model states generated by running the model with an ensemble of meteorological forcing inputs. Because we do not perturb individual model states, the water balance remains consistent before and after resampling. Few hydrologic studies on data assimilation have included realistic meteorological forcing errors, and those that do typically employ satellite

precipitation error models [e.g., *Moradkhani et al.*, 2006] that are unlikely to describe the error characteristics of forcings used in operational streamflow prediction practice.

To our knowledge, this is the first hydrologic modeling paper that uses a particle filter in which errors in input precipitation and temperature are derived empirically from a probabilistic analysis of meteorological observations. Probabilistic interpolation of the station data was performed using spatial regression methods [*Clark and Slater,* 2006; *Newman et al.,* 2015] to generate an ensemble of equiprobable gridded precipitation and temperature daily time series. This approach provides quantitative uncertainty estimates for the meteorological fields that can be leveraged in the DA. Ensemble spread represents uncertainty in the raw observations and in their spatial interpolation.

Figure 2.1 outlines the general workflow used for simulating the ensembles of model states that represent the uncertainty in the hydrologic model attributable to errors in meteorological forcing data. In situ observations of daily precipitation and temperature were inputs to a spatial regression approach [*Newman et al.,* 2015] used to create N_p equally probable daily meteorological model input time series (section 2.3.1), each of which was then used to force a hydrologic and channel routing model (section 2.3.2) to simulate daily hydrologic conditions. Daily streamflow and model states were saved for each ensemble member for use by an external DA analysis module.

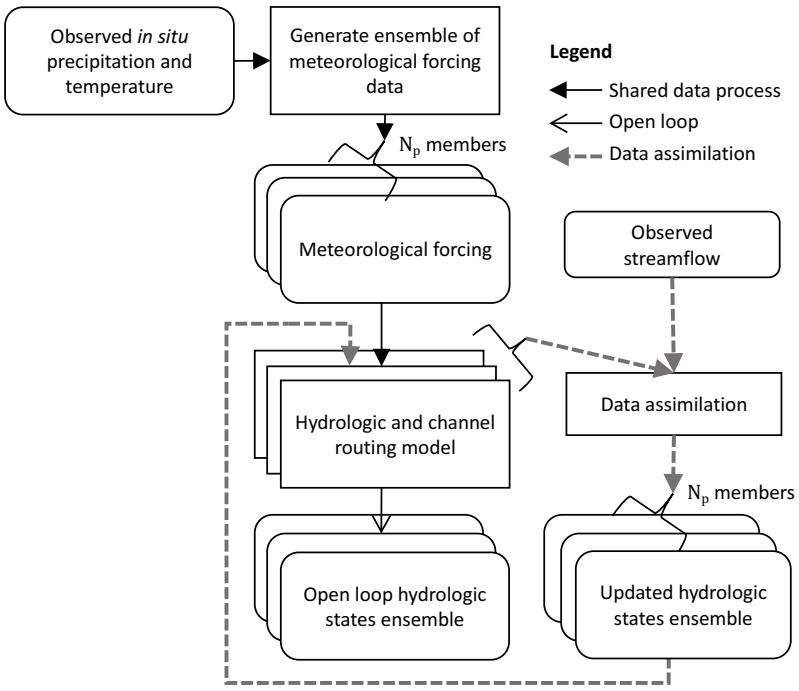

Figure 2.1. Workflow to generate ensemble time series of hydrologic model states for open loop and data assimilation cases. In this study, observed streamflow is synthetic.

The daily streamflow and model states from a single continuous simulation for each ensemble member constitute the open loop hydrologic conditions. In addition, a set of five DA experiments was conducted using different particle filter implementations (sections 2.2.2 and 2.2.3). In all DA experiments, the hydrologic model was used to propagate hydrologic conditions for each ensemble member until an observation was available (essentially daily at our sites). Once an observation of streamflow becomes available, the DA technique updates the ensemble states (Fig. 2.1) that are then propagated through the hydrologic model until the next observation is available, at which point the states are once again updated.

2.2.2 *Particle filters*

Particle filters are attractive for hydrologic modeling applications because they can be applied to nonlinear and non-Gaussian problems. We start with the forward model:

$$x_t = f_t(x_{t-1}, v_{t-1}) \tag{2.1}$$

where x_t is the model state at time t, and f_t is a function of x_{t-1} (the model state at time $t-1$) and of v_{t-1} (the process noise at time $t-1$). Each of these elements can be either scalar or vector quantities. For most hydrologic models, the model states include soil moisture and where relevant, snow storage (snow water equivalent). The process noise is any type of uncertainty, such as uncertainty in model parameters, atmospheric forcings, and model physics. Only errors in forcing data are explicitly represented in the implementation presented here – thus, the process noise, v_t, is incorporated by running the forward model with an ensemble of meteorological forcings.

The goal of the particle filter is to estimate a probability distribution of simulated states, given an uncertain observation z_t,

$$z_t = h_t(x_t, n_t) \tag{2.2}$$

where h_t is a function that relates the model state to the observation (known as the "observation operator"), and n_t is the measurement error at time t. In this case, the measurements are observations of streamflow, but the same methodology could be applied to observations of additional hydrologic states and fluxes. In this paper, n_t is specified using a Gaussian function with mean x_t and standard deviation specified based on published streamflow gaging error estimates. As with Eqn. 2.1, the elements of Eqn. 2.2 can be scalar or vector quantities.

Particles (e.g., hydrologic simulation ensemble members) are assigned weights proportional to each particle's likelihood given the observations [*Kitagawa*, 1996] and

observational error. The particle weights define the posterior probability distribution function ($p_t(x)$ at time t) for model states x_t given the observations $z_{1:t}$:

$$p_t(x_t|z_{1:t}) \approx \sum_{i=1}^{N_p} w_t^i \, \delta\left(x_t - x_t^i\right) \tag{2.3}$$

where w_t^i and x_t^i are the normalized weight and state of particle i, respectively, for a total of N_p particles, and $\delta(\cdot)$ is the Dirac delta function [*Kitagawa*, 1996; *Doucet et al.,* 2001; *Arulampalam et al.*, 2002].

The general particle filter workflow (Fig. 2.2) starts with a series of N_p equally weighted particles, which we sample from the retrospective hydrologic conditions ensemble at time t_0. Each particle is run forward, with the same weight, until the next time step for which there is an observation. At this point, new particle weights are calculated. The weighting functions used in this study are described in section 2.2.3. As the ensemble evolves, some particles may diverge significantly from the observations, in which case they are considered ineffective. When there are too few effective particles, the ineffective particles are discarded and the remaining particles are resampled and given new weights, restoring N_p effective particles in the filter (see section 2.2.2). The hydrologic model then propagates the particle states forward in time until the next observation or until the time at which a forecast will be initialized. How the states are used to initialize the forecast is outside of the scope of this study.

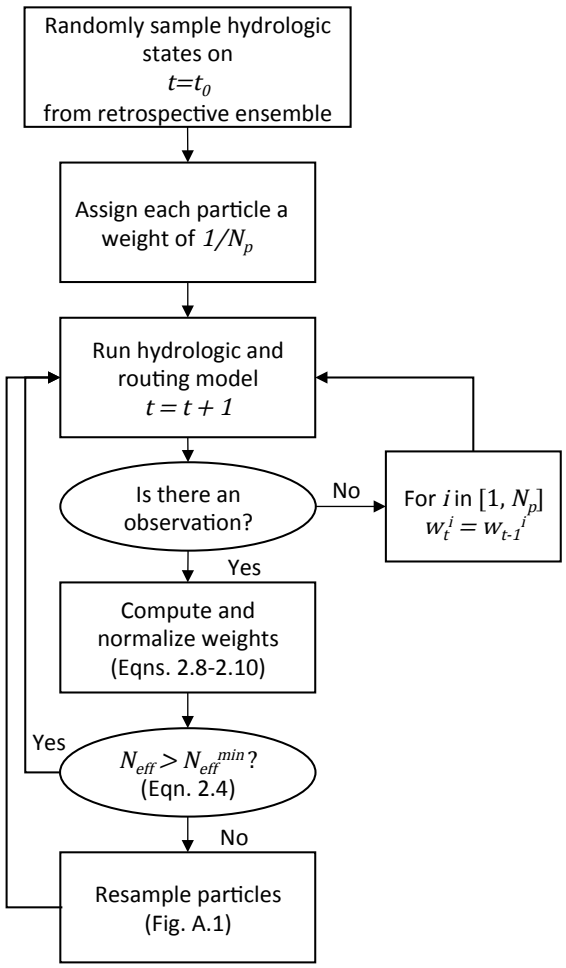

Figure 2.2**.** Particle filter flow chart.

Two well-known issues in the application of particle filters are 1) sample degeneracy, which occurs when the weights of very few particles dominate, and its consequence, 2) sample impoverishment, which occurs when the filter only propagates these dominant particles into the future, rather than the entire distribution [*Li et al.*, 2014]. Because of the dependency of weights in sequential schemes on their value at previous time steps, this problem is particularly notable in methods like Sequential Importance Sampling, but occurs in non-sequential particle filters as well. Resampling schemes have been developed to minimize sampling degeneracy; however, they do not necessarily combat sample impoverishment [*Salamon and Feyen*, 2010]. The number

of effective particles, N_{eff}, is used as a measure of sampling degeneracy [*Kong et al.*, 1994; *Liu and Chen*, 1995]. N_{eff} can be estimated as follows [*Doucet et al*., 1999]:

$$N_{eff} = \frac{1}{\sum_{i=1}^{N_p} w_t^{i\,2}} \tag{2.4}$$

In this paper, we selected $N_p/5$ as the minimum number of effective particles required. Whenever N_{eff} falls below this minimum, the particle distribution is resampled.

The resampling algorithm used here is based on the resampling algorithm presented in *Arulampalam et al.* [2002], *Moradkhani et al.* [2005], and *Weerts and El Serafy* [2006]. The primary steps of the resampling scheme are outlined in Fig. A.1 (Appendix A). The algorithm creates a set of equally weighted particles that reflect the probability distribution from the particle weights prior to resampling.

2.2.3 *Experiments*

2.2.3.1 Open loop and synthetic truth

In our experiments, we use 99 particles ($N_p = 99$) plus one truth simulation. Each of 100 meteorological ensemble members is used to force each of 100 hydrologic model simulations. As described in section 2.2.1, these meteorological ensemble members are equiprobable by design, so the distribution of the equally weighted resultant hydrologic states on each day represents that day's prior distribution of IHCs. One member of the ensemble is taken as a synthetic "truth" for the entire duration of the simulation, and the remaining 99 members are used as the open loop (no DA) case.

2.2.3.2 Particle Filters

As in *Moradkhani et al.* [2005], we use a likelihood function, $p(z_t|x_t^i)$, at time, t, for particle, i, that is a Gaussian function of the difference between the observed value, z_t, and the modeled value, x_t^i, with a standard deviation of σ_t:

$$p(z_t|x_t^i) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{(z_t - x_t^i)^2}{2\sigma_t^2}} \tag{2.5}$$

We normalize the values of $p(z_t|x_t^i)$ to sum to unity before we use them in weight calculations. The standard deviation, σ_t, in Eqn. 2.5 represents the observational uncertainty. Eqn. 2.5 derives from an assumption that n_t in Eqn. 2.2 fits a Gaussian error distribution, $N(0, \sigma_t^2)$. We use a σ_t value of 10% of the observed flow at time t, given that the U.S. Geological Survey (USGS) National Water Information System reports streamflow data quality from poor (>8% of the actual flow) to excellent (within 2% of the actual flow). The streamflow estimates in our study represent reservoir inflows that are less well measured than many streamflow locations. To assess the sensitivity of this choice, we also conduct experiments with a σ_t value of 25% of the observed flow.

Doucet et al. [2001] present a discussion of Sequential Importance Sampling, which results in the following widely used equation for particle weights, in which the weight from the prior assimilation time is multiplied by the weight of the current assimilation time:

$$w_t^i \propto w_{t-1}^i \cdot p(z_t|x_t^i) \tag{2.6}$$

Weights are normalized to sum to unity. This method is prone to sample degeneracy [*Li et al.*, 2014]. Furthermore, this recursive behavior complicates the interpretation of particle weights on any given day because sporadic poor behavior that occurred several or many days prior could

result in low weights for a particle that otherwise performs well. That is, particles are heavily penalized for past poor behavior even though they may perform well for current conditions.

To assess the impact of the long performance memory inherent in the recursive weight approach, we compare it with a simpler approach that has limited performance memory. Instead of calculating weights recursively, we calculate the weight each day as a function of model performance over a series of prior days, with more weight given to recent model performance. In this case,

$$w_t^i \propto \sum_{j=t-n_{wt}}^{t} p\left(z_t \middle| x_t^i\right) e^{[-c(t-j)]} \tag{2.7}$$

where n_{wt} is the number of days prior to t that are included in the calculation of w_t^i, c is an exponential decay constant that determines the degree of influence of each day on w_t^i. This essentially limits the performance memory to this resampling window of n_{wt} days. The exponential decay multiplier is shown in Fig. 2.3 for c values of 0, 0.5, 1, and 25 day^{-1}. Note that the $p\left(z_t \middle| x_t^i\right)$ for all 7 days is weighted equally if c equals 0, and only $p\left(z_t \middle| x_t^i\right)$ on the day prior to forecast is considered if c is very large (c =25 in this example).

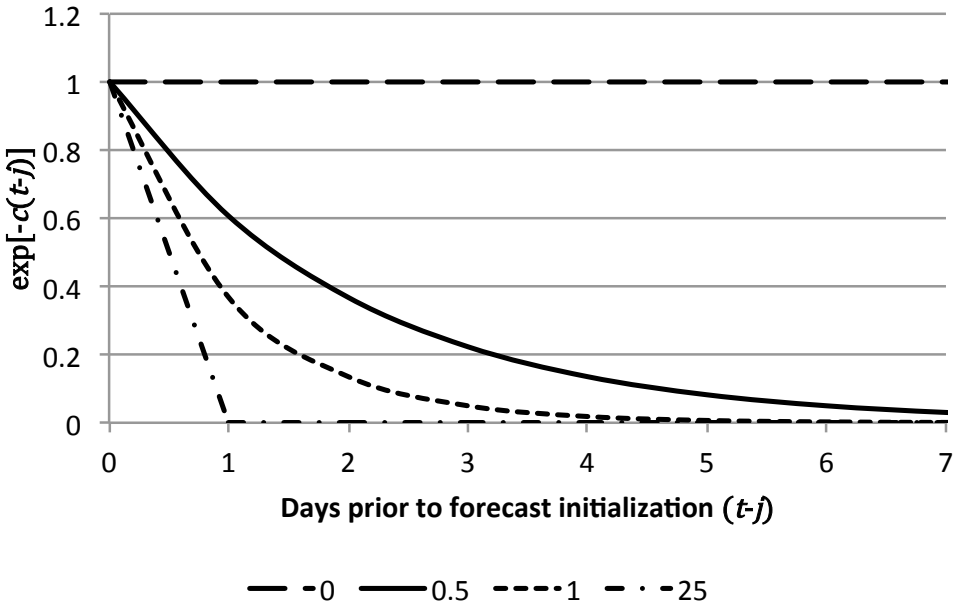

Figure 2.3. Dependence of decay function used in Eqn. 2.7 on choice of c. For larger c, performance on only the day prior to the forecast is used in weighting particles; while days further in the future have a larger impact on the weights with lower c.

We can combine the likelihood function in Eqn. 2.5 with the weight functions in Eqns. 2.6 and 2.7 to form the equations used to calculate weights for each particle filter implementation. Substituting Eqn. 2.5 into Eqn. 2.6 gives the proportional (non-normalized) weights, w_t^{i*}, for the sequential Gaussian particle filter:

$$w_t^{i*} = w_{t-1}^{i} \cdot \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{\left(z_t - x_t^i\right)^2}{2\sigma_t^2}} \cdot \left(\sum_{m=1}^{N_p} \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{\left(z_t - x_t^m\right)^2}{2\sigma_t^2}} \right)^{-1} \tag{2.8}$$

Substituting the same likelihood function (Eqn. 2.5) into Eqn. 2.7 results in the equation for proportional (non-normalized) weights used in the non-sequential Gaussian particle filter:

$$w_t^{i*} = \sum_{j=t-n_{wt}}^{t} \left[e^{-c(t-j)} \cdot \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{(z_t - x_t^i)^2}{2\sigma_t^2}} \cdot \left(\sum_{m=1}^{N_p} \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{\left(z_t - x_t^m\right)^2}{2\sigma_t^2}} \right)^{-1} \right] \tag{2.9}$$

In order to represent the probability distribution as defined in Eqn. 2.3, weights calculated in Eqns. 2.8-2.9 are normalized to sum to one as:

$$w_t^i = \frac{w_t^{i*}}{\sum_{m=1}^{Np} w_t^{m*}} \tag{2.10}$$

2.2.4 *Evaluation of ensembles*

We use the continuous rank probability score (CRPS) [*Hersbach*, 2000] to compare the performance of the different PF methods. The CRPS is a widely-used metric for evaluating how well ensembles represent single-valued observations. The CRPS compares the cumulative probability distribution (CDF) of each simulated state s on day t, $F_t^s(s)$, with that day's "true" CDF (based on on observations), $F_t^y(s)$, where the superscripts s and y denote simulated and truth, respectively. This value is calculated over n days as:

$$\text{CRPS} = \frac{1}{n}\sum_{t=1}^{n}\int_{s=-\infty}^{\infty}\left(F_t^s(s) - F_t^y(s)\right)^2 ds \tag{2.11}$$

CRPS assumes that the observations are perfect, such that $F_t^y(s)$ is defined as the Heaviside function conditioned on the observed hydrologic states, y_t:

$$F_t^y(s) = \begin{cases} 0, & s < y_t \\ 1, & s \geq y_t \end{cases} \tag{2.12}$$

CRPSS penalizes distributions for bias and a large spread. It essentially measures the area between the simulated and observed distributions. Distributions that best match the observations will have CRPS values close to zero.

To facilitate comparison across study sites and hydrologic variables, CRPS is converted to a skill score using the CRPS of a reference distribution, CRPS_{ref}:

$$\text{CRPSS} = 1 - \frac{\text{CRPS}}{\text{CRPS}_{ref}} \tag{2.13}$$

In this study, we use the CRPS of the open loop case as CRPS_{ref}. Estimated IHC distributions with high probabilistic skill will have CRPSS close to one. If CRPSS is less than or equal to zero, the open loop distribution performs better than the particle filter. Results presented here are calculated in Python using the properscoring library [*The Climate Corporation,* 2015].

2.3 CASE STUDIES

We conducted a series of experiments using watershed simulations of inflows to two reservoirs in the Pacific Northwest (PNW) region in the United States (Fig. 2.4; Table 2.1). Hungry Horse reservoir is located on the South Fork Flathead River in Montana, and Howard Hanson Dam is located on the Green River in Washington State. Both basins are heavily forested, with forest making up 83% of the area above Hungry Horse [*McCarthy et al.*, 2016] and 91% of the drainage area above Howard Hanson [*Sumioka et al*., 1998]. Both are headwater basins with minimal regulation above the respective reservoirs. Two primary differences in hydrologic response of these basins are the form of precipitation and the time of concentration (time to travel the hydraulically longest path from the basin boundary to the outlet). Precipitation in the basin above Hungry Horse, as in much of the PNW, is winter-dominant and falls primarily as snow. Above Howard Hanson, precipitation falls as a mixture of rain and snow with most of the runoff generated during fall (October through December) rain events. The time of concentration, as defined by the calibrated unit hydrograph, is on the order of 3.5 days at Hungry Horse, while the time of concentration at Howard Hanson is closer to 0.9 days.

Table 2.1. Physical properties of Green River Basin upstream of Howard Hanson Dam, and the South Fork Flathead River upstream of Hungry Horse Dam. Hydrologic variables correspond to basin-averaged means for the period October 1980 to September 2015. Precipitation is the mean of the gridded ensemble. Runoff is the simulated mean of the retrospective hydrologic ensemble. Potential evapotranspiration (PE) was calculated using the Priestley-Taylor method.

	Howard Hanson	**Hungry Horse**
Area (km²)	570	4200
Basin average elevation (m.a.s.l.)	905	1773
Mean annual precipitation, P (mm/yr)	1890	1043
Simulated mean annual runoff, R (mm/yr)	1483	676
Simulated mean annual runoff ratio (R/P)	0.78	0.65
Simulated mean annual dryness index (PE/P)	0.63	1.22

2.3.1 *Meteorological Data*

In situ station observations of daily minimum and maximum temperature and precipitation were obtained from the Daily Global Historical Climatology Network and updated daily (GHNC-Daily) [*Menne et al.,* 2012a; 2012b]. For our study areas, this data set includes stations from the Community Collaborative Rain, Hail, and Snow network (CoCoRaHs), U.S. Natural Resources Conservation Service SNOwpack TELemetry (SNOTEL), the U.S. Cooperative Network, and Environment Canada.

These station observations were interpolated following a spatial regression approach [*Clark and Slater,* 2006; *Newman et al.,* 2015] to generate an ensemble of 100 realizations of equiprobable gridded precipitation and temperature daily time series at 1/16th degree latitude by longitude grid resolution. Each member of the gridded ensemble time series was areally averaged to the hydrologic response units for each basin using the Poly2Poly Python-based conservative regridding tools developed at NCAR (https://www.github.com/NCAR/Poly2Poly.git). One week

of the basin-averaged daily meteorological forcing time series ensemble is shown in Fig. 2.4. Because station data poorly constrain estimates of spatially continuous (i.e., gridded) meteorological fields, the estimation process can result in vastly different behavior between ensemble members on any given day. For example, on 4 Jan. 2010, the ensemble at Hungry Horse reports precipitation between 0 and 45 mm and at Howard Hanson between 0 and 65 mm. Temperature estimates are more certain; however, the ensemble still shows a notable spread. Such results support the argument that additional constraints on basin hydrology, such as assimilation of streamflow observations, are needed to simulate realistic model states.

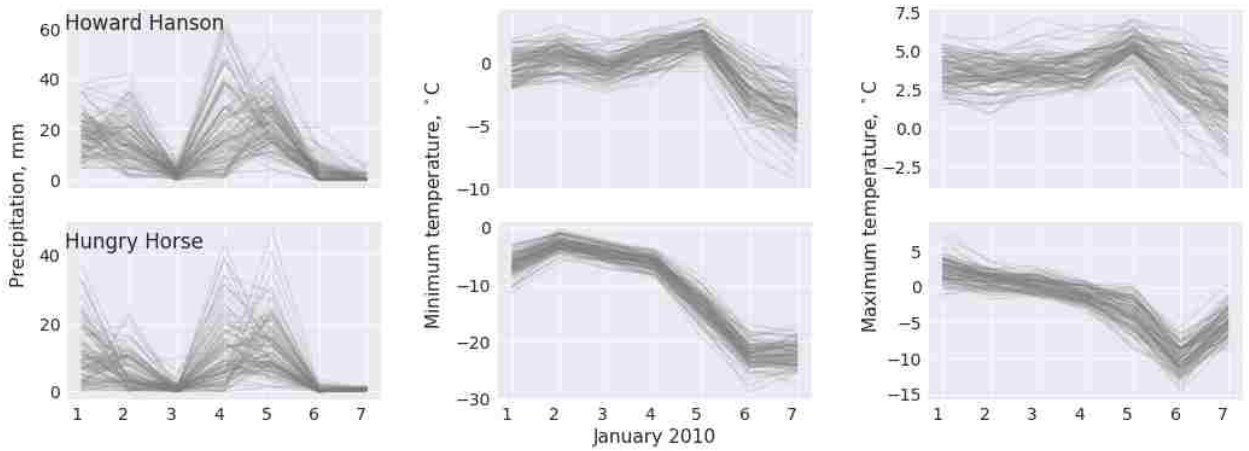

Figure 2.4. An example of daily time series of precipitation, and minimum and maximum temperature at the Howard Hanson and Hungry Horse basins for 100 equally probable ensemble members. Forcings for the first week in January 2010 are shown.

2.3.2 *Hydrologic modeling*

For the hydrologic and channel routing model, we used the Sacramento Soil Moisture Accounting model (SAC-SMA) [*Burnash et al.,* 1973] and the Snow-17 model for snow accumulation and ablation [*Anderson*, 1973], with unit hydrograph flow routing. This combined modeling approach was first developed by the National Weather Service (NWS) and will be

referred to as the NWS models for the rest of this paper. Each study basin was simulated with two elevation bands, and the model parameters were calibrated for this study for the years 1981-2008. The calibration used a forcing time series constructed by randomly selecting year-long segments from a different meteorological forcing ensemble member. This process was found to reduce model bias when the model was run with the entire forcing ensemble. The Multi-Objective Complex evolution Optimization Model (MOCOM) [*Yapo et al*., 1998] was used to calibrate the model by minimizing two objective functions: average root-mean-square error in long-term monthly means and root-mean-square error of daily flows, relative to naturalized streamflow observations. Nash-Sutcliffe Efficiency scores of 0.70 and 0.86 for daily streamflow values were obtained for the Howard Hanson and Hungry Horse basin models, respectively, reflecting strong but not outstanding calibration quality. Observed naturalized daily flows from 1980-2008 were obtained from the No-Regulation No-Irrigation (NRNI) streamflow dataset developed by the Columbia River Management Joint Operating Committee [*CRMJOC*, 2015] (Figs. A.2-A.3; Appendix A). A ten-year model simulation period was used in the DA assessment, spanning 3,652 days between 1 October 2005 and 30 September 2015.

2.4 RESULTS

CRPSS values, calculated from daily streamflow, soil moisture and SWE, respectively, for water years 2007-2015, are shown in Fig. 2.5 for $\sigma_t = 0.10z_t$ (more certain observations) and in Fig. 2.6 for $\sigma_t = 0.25z_t$ (less certain observations). For both sites, all particle filters improve the ensemble of simulated streamflow relative to the open loop simulations, and the sequential particle filter produces the highest CRPSS values for streamflow. The open loop case produces a more statistically reliable streamflow distribution than the filter in only two cases: 1) for $\sigma_t = 0.10z_t$ with the intermediate memory non-sequential filter (c = 1) in water year 2012 at Hungry

Horse (Fig. 2.5), and 2) for $\sigma_t = 0.25z_t$ with the long memory non-sequential filter ($c = 0$) in all water years at Howard Hanson (Fig. 2.6). As expected, all particle filters show a greater improvement in CRPSS for streamflow when smaller errors are assumed for the observations. Higher certainty in observations, however, does not translate into higher CRPSS for soil moisture and SWE in many cases. The performance of the particle filters varies by year.

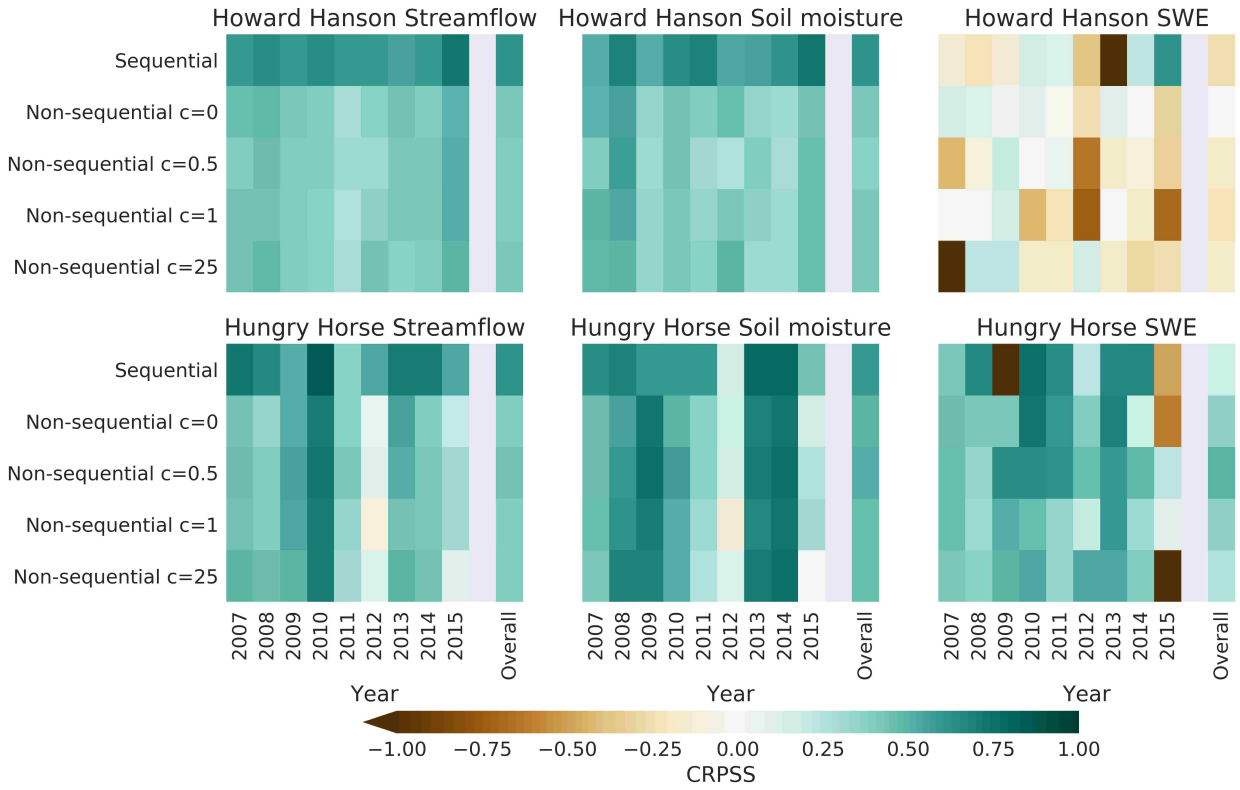

Figure 2.5. CRPSS values for Howard Hanson (top row) and Hungry Horse (bottom row) for each simulation by year and overall, for different particle filters with σ_k=0.10z_t. Skill scores are calculated using the CRPS of the open loop case for the same period. Scores are shown for streamflow (left), soil moisture (center) and SWE (right).

Figure 2.6. CRPSS values for Howard Hanson (top row) and Hungry Horse (bottom row) for each simulation by year and overall, for different particle filters with σ_k=0.25z_t. Skill scores are calculated using the CRPS of the open loop case for the same period. Scores are shown for streamflow (left), soil moisture (center) and SWE (right). Note the color scale differs from that in Fig. 2.5.

As with streamflow, the sequential filter produces better overall CRPSS for soil moisture than the non-sequential filters at both sites and for each level of assumed error in the observations. At Howard Hanson, all particle filters improve the CRPSS for soil moisture in all years when smaller observation error is used (Fig. 2.5), and all particle filters with higher observational error assumptions show minimal improvement in overall CRPSS (Fig. 2.6). While the overall CRPSS is higher for soil moisture at Howard Hanson for $\sigma_k = 0.10z_t$, the highest CRPSS of soil moisture in 2015 is produced by non-sequential particle filters for $\sigma_k = 0.25z_t$. Water year 2015 is one of the driest years of this period, and there are virtually no peaks in

streamflow after mid-April. Starting in mid-March 2015, the open loop case is biased high in streamflow and soil moisture (and SWE starting in January), suggesting that most of the meteorological ensemble members have more precipitation than actually happened during this period. Because the sequential filter matches streamflow very closely up until this point in time, it has not evolved states that have room to store the additional precipitation. The non-sequential filter maintains a larger spread in model states, and because it only accounts for the recent past in its weights, more readily drops those particles that could not absorb the excess precipitation. As the season progresses, there is very little streamflow, so there are no opportunities for the filters to further adjust, resulting in persisting bias in the sequential filter.

At Hungry Horse, the open loop case never outperforms the filters in terms of CRPS of soil moisture when observations are assigned error of $0.25z_t$ (Fig. 2.6) for all particle filters; for $\sigma_k = 0.10z_t$, the CRPS of soil moisture for the open loop case is slightly better than that of the non-sequential ($c = 1$) particle filter in 2012 only.

For SWE, on the other hand, the sequential filter only outperforms the non-sequential filters when the streamflow observations have more error (Fig. 2.6). When the observations are assigned less error (Fig. 2.5), a non-sequential filter with intermediate memory ($c = 0.5$) performs best at Hungry Horse, and the open loop case and the long-memory non-sequential filter ($c = 0$) perform best at Howard Hanson. Regardless of the assumed observational error, the open loop case performs better, or as well as, most of the filters in most years at Howard Hanson in terms of CRPS for SWE. At Hungry Horse, the open loop case only outperforms three of the filters in 2015, and the sequential filter in 2009, if the observations are assigned less error.

To gain insight into the behavior of the filters, daily time series showing the evolution of particle states and streamflow distributions during water year 2007 are plotted in Figs. 2.7-2.12

for $\sigma_k = 0.10z_t$, which outperformed the $\sigma_k = 0.25z_t$ filter configuration. The time series in Figs. 2.7-2.12 compare the ensemble distribution quantiles with the synthetic observation on each day of analysis. Because of resampling, the particle distribution on any given day may have evolved from only a subset of the particle distributions leading up to that day. For example, for the case of c = 25 at Howard Hanson in Fig. 2.9, the high springtime SWE values are not included in any of the time series for the particles that remain by Aug. 1, 2007 because the particles corresponding to high springtime SWE were discarded in late May/early June.

For water year 2007, which is representative of the overall CRPSS results at Howard Hanson, all particle filters succeed in reducing the magnitude of the first peak in November in the open loop simulations, as well as a smaller event in January, to better match observations (Fig. 2.7). The open loop ensemble median soil moisture is lower than the (synthetic) true soil moisture at the beginning of the water year (Oct. through Jan.; Fig. 2.8). All of the filters remove this bias (Fig. 2.8). For SWE, however, the shorter memory non-sequential particle filters accumulate more snow in the early (c = 1) or late season (c = 0.5 and c = 25) in 2007 than in the synthetic truth (Fig. 2.9). The median SWE simulated with the longer memory non-sequential filter (c = 0) and the sequential filter track observed SWE better in this case, but the spread of particles in the sequential filter does not contain the observed SWE for much of the mid-winter (Fig. 2.9).

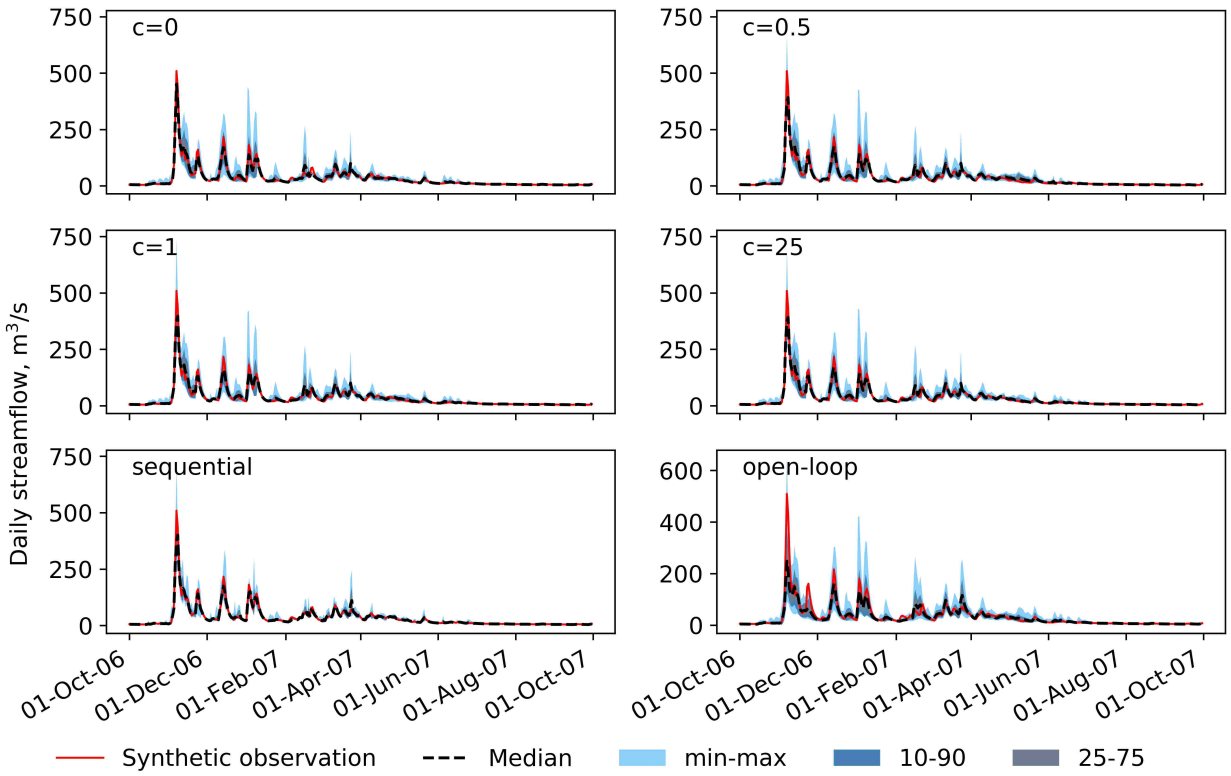

Figure 2.7. Daily streamflow time series at Howard Hanson for water year 2007, for particle filter runs with $\sigma_k = 0.10 z_t$. Colored bands show minimum and maximum ensemble values, the 10th to 90th percentile, and the 25th to 75th percentile.

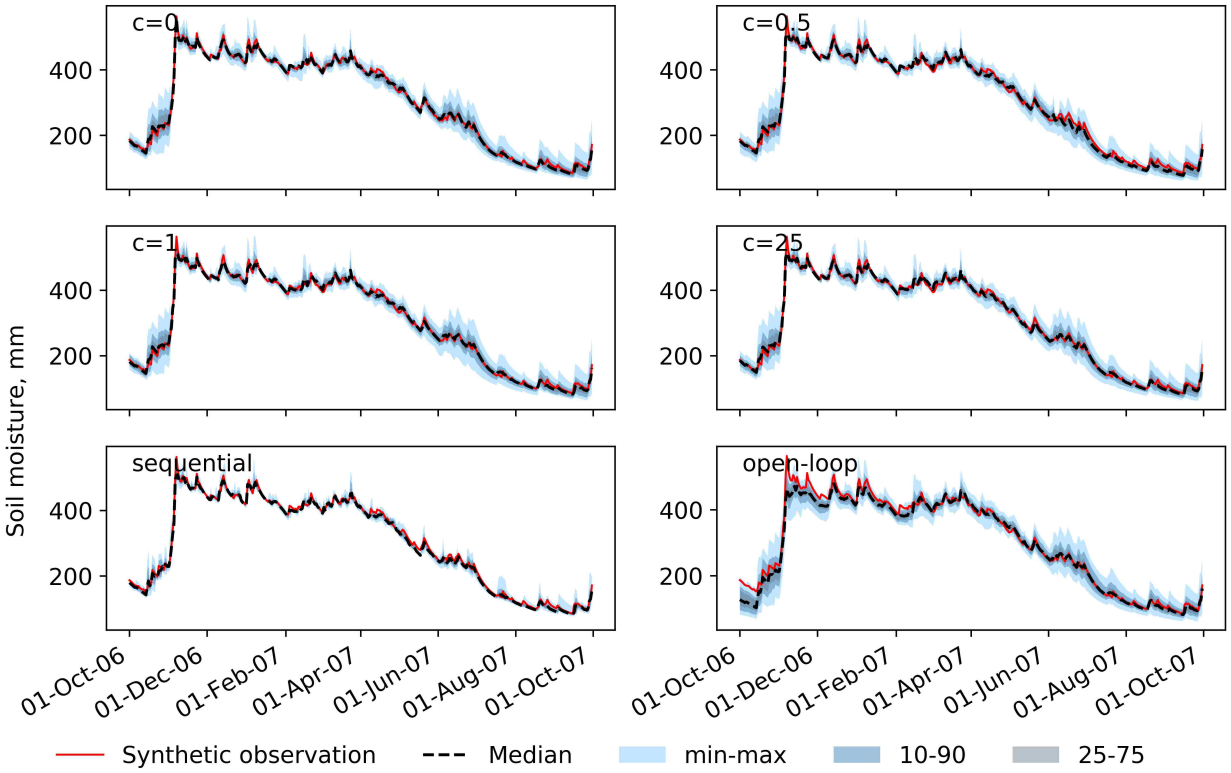

Figure 2.8. Daily soil moisture time series at Howard Hanson for water year 2007, for particle filter runs with $\sigma_k = 0.10 z_t$. Colored bands show minimum and maximum ensemble values, the 10th to 90th percentile, and the 25th to 75th percentile.

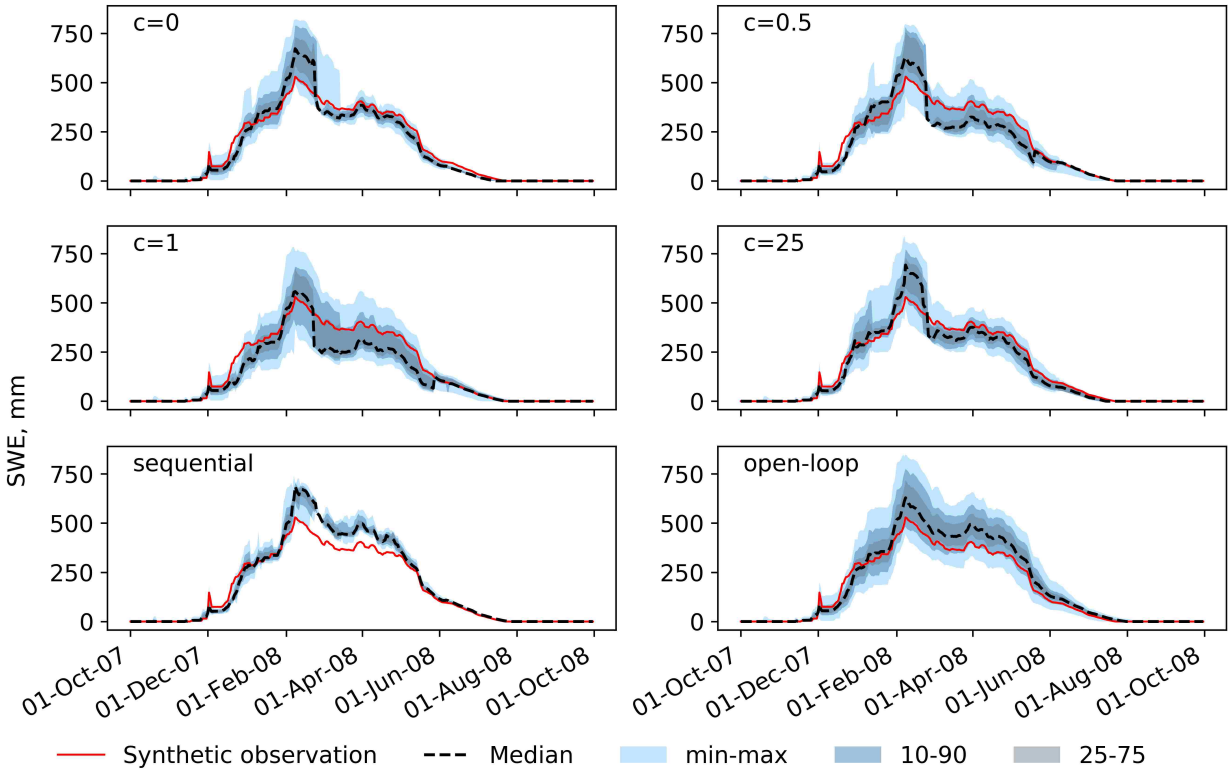

Figure 2.9. Daily SWE time series at Howard Hanson for water year 2007, for particle filter runs with σ_k=0.10z_t. Colored bands show minimum and maximum ensemble values, the 10th to 90th percentile, and the 25th to 75th percentile.

For water year 2013, which reflects the overall CRPSS pattern at Hungry Horse, the open loop ensemble median of streamflow matches the synthetic observation well overall, but misses the magnitude of the November peak entirely, as well as two smaller events in late January (Fig. 2.10). For the open loop ensemble, the soil moisture during and immediately prior to these events is also lower than the synthetic true soil moisture (Fig. 2.11). Because the open loop SWE is higher than the synthetic true SWE during the November and January events, at least part of the discrepancy in streamflow is due to excess snow accumulation in the open loop case. Applying any of the particle filters improves SWE because the streamflow events are closely tied to snowmelt. For example, to match higher streamflow in November, the particle filter favors

particles with lower snow accumulation (or higher melt) during that event (Fig. 2.12). The longer memory filters (sequential and non-sequential c = 0) produce a better CRPSS for all variables in this basin (Fig. 2.5). This is because they produce a smaller spread than the shorter memory cases, and because the time of concentration for Hungry Horse is multiple (~3.5) days. When the observational errors are lower, the CRPSS is uniformly high relative to the open loop case for SWE for all filters because the range of simulations is too high to discard most particles.

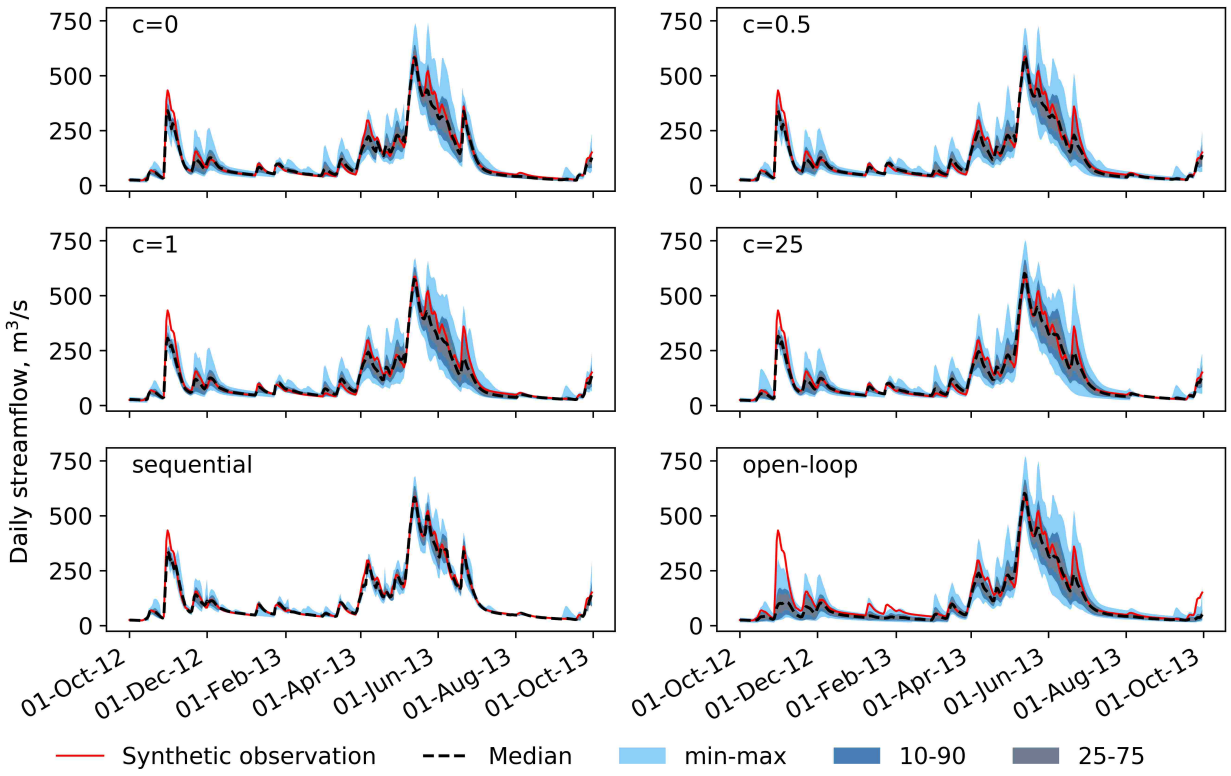

Figure 2.10. Daily streamflow time series at Hungry Horse for water year 2013, for particle filter runs with σ_k=0.10z_t. Colored bands shows minimum and maximum ensemble values, the 10th to 90th percentile, and the 25th to 75th percentile.

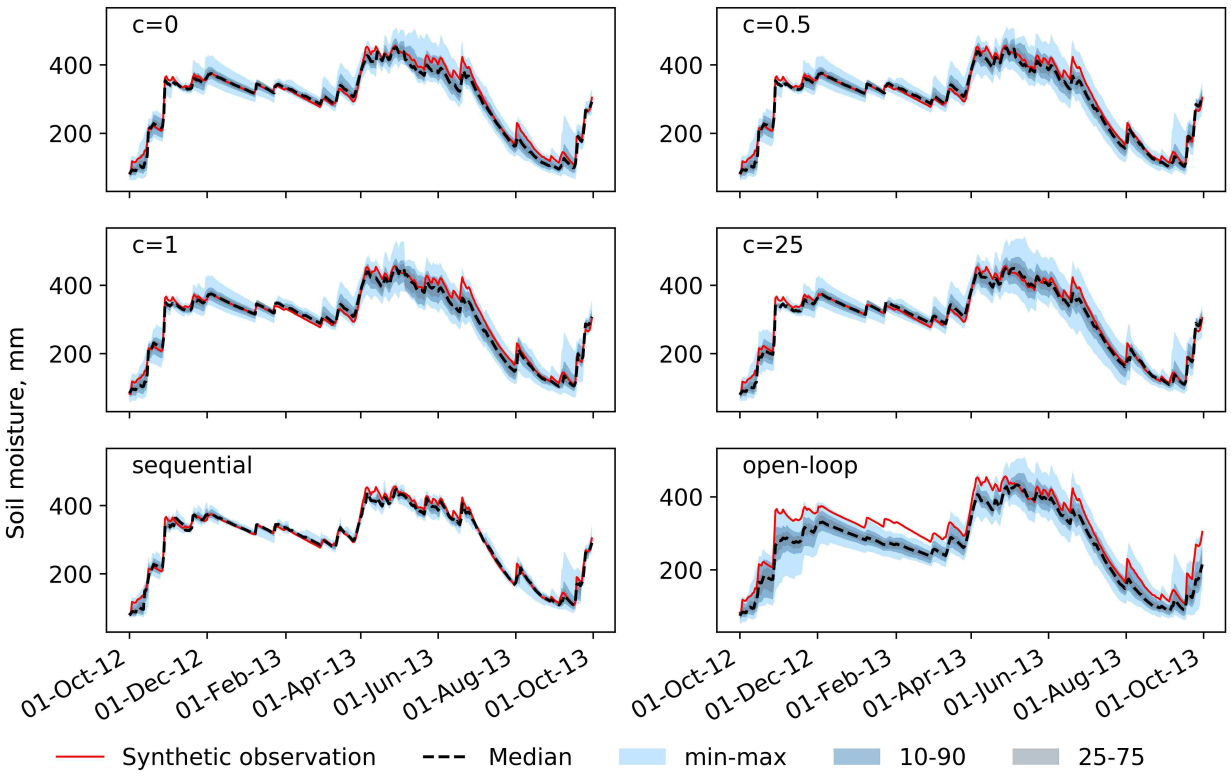

Figure 2.11. Daily soil moisture time series at Hungry Horse for water year 2013, for particle filter runs with $\sigma_k = 0.10 z_t$. Colored bands shows minimum and maximum ensemble values, the 10^{th} to 90^{th} percentile, and the 25^{th} to 75^{th} percentile.

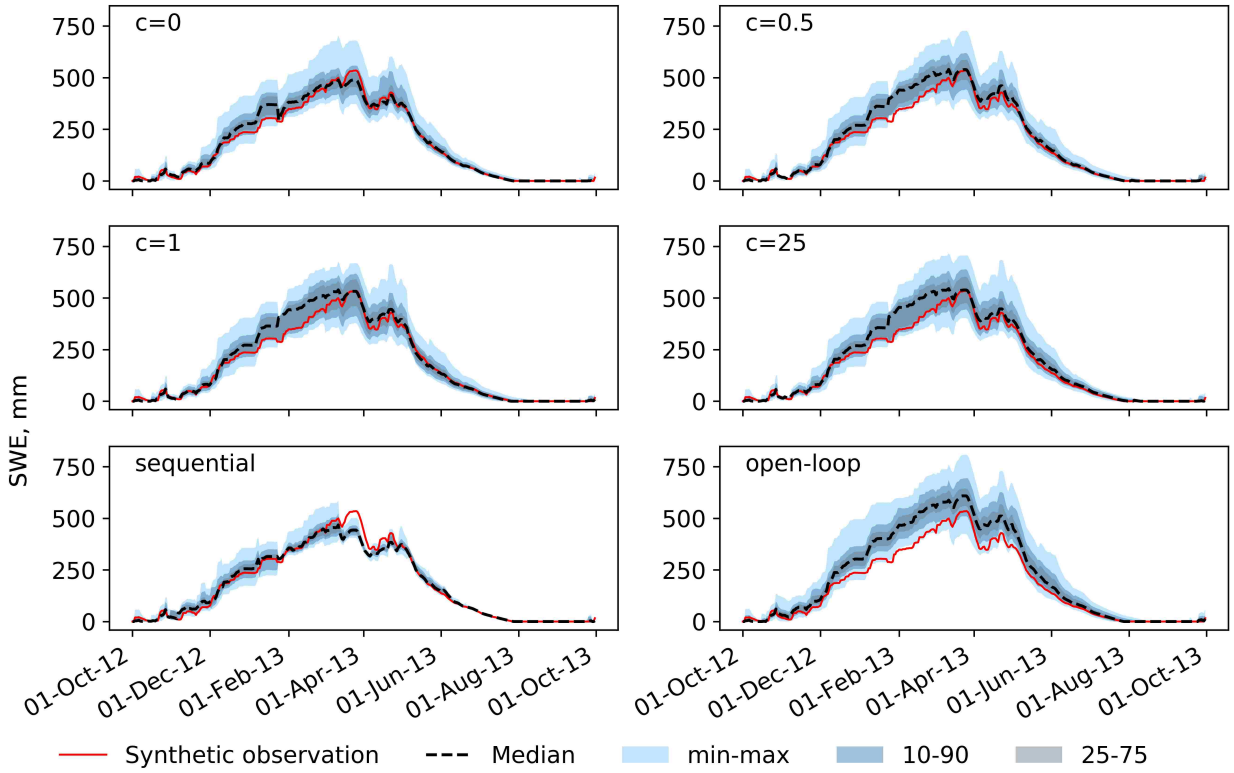

Figure 2.12. Daily SWE time series at Hungry Horse for water year 2013, for particle filter runs with $\sigma_k = 0.10 z_t$. Colored bands shows minimum and maximum ensemble values, the 10th to 90th percentile, and the 25th to 75th percentile.

2.5 Discussion

Even though the sequential filter outperforms the non-sequential filter in terms of CRPSS, the non-sequential filter improves upon the open loop case, and in some cases performs almost as well as the sequential filter. In nearly all cases, higher CRPSS in the sequential case can be attributed to a smaller ensemble spread. This small spread results directly from the recursive application of weights.

The filter behavior at Hungry Horse differs from the behavior at Howard Hanson largely because snow accumulation and melt play a more important role in the timing of streamflow. At certain times of year, the streamflow timing and magnitude are more reflective of snow storage

at Hungry Horse than at Howard Hanson. Because most of the variability in streamflow at Howard Hanson is driven by rainfall events, the streamflow observations contain less information about snow storage.

Although the time of concentration at Hungry Horse is 3 times that at Howard Hanson, both basins have short times of concentration relative to the daily model time step. As a result, all non-sequential filters perform similarly in terms of CRPSS of streamflow and soil moisture at both sites (Figs. 2.5-2.6). Because the time of concentration at Howard Hanson is less than one day, previous streamflow over extended periods adds little to and even degrades the estimation of current streamflow. As a result, if the observations are given lower weight in the assimilation ($\sigma_t = 0.25z_t$), equally weighting the past 7 days produces a less representative distribution of streamflow overall than using only the current day's streamflow (Fig. 2.6).

If streamflow observations are given lower weight, the particle filter maintains a larger spread of streamflow ensemble members, reflecting greater emphasis on simulated states. The CRPSS is penalized for wider spread when observations have more error, but the wider spread allows for a broad range of model simulations to continue even if they miss the observed flow on a given day. This range is important if the observed flow is inaccurate or contains limited information about individual hydrologic states, such as non-melting SWE or same-day soil moisture, which may influence runoff only at a lag. When streamflow observations are given more weight (lower error), the particle filters do not retain as many particles, and as a result, the particle filters are not able to recover if states such as SWE state diverge. For this reason, the CRPSS for SWE is worse in some years at both sites for $\sigma_t = 0.10z_t$ (Fig. 2.5) than for $\sigma_t = 0.25z_t$ (Fig. 2.6).

Snyder et al. [2008] emphasize that resampling does not improve the current posterior distribution, but rather hopes to prevent future degeneracy by propagating more likely states. They demonstrate that sample degeneracy becomes increasingly problematic as the number of states grows. In the case of lumped models, like the NWS models used here, the number of states is constrained by using a small number of hydrologic response units; however, in hydrologic applications that use distributed models, the number of spatially distributed states can be quite large [*Salamon and Feyen*, 2010]. The non-sequential filters presented here require resampling approximately only 25% as frequently as the sequential filter at either site. This means that the sample is less likely to degenerate for the non-sequential filter. This may be more valuable in real-world applications, in which model bias can cause the distribution of simulations to diverge farther from the observations than in this synthetic case.

2.6 CONCLUSIONS

Using synthetic data assimilation experiments and hydrologic simulations driven by meteorological ensemble forcings, we have assessed the performance of a recursive (SIR) particle filter methods in improving simulation performance. We have evaluated the impact of the long memory assumptions of this filter in contrast to a non-sequential particle filter implementation with more direct control over the representation of system memory. We tested this framework with four representations of system memory against an open loop and sequential particle filter implementation. We also examined the sensitivity of these implementations to the error assigned to streamflow observations. The results of these experiments highlight the complexity of interactions between streamflow and hydrologic states, showing that the particle filter that produces the best estimates of prior streamflow does not always produce the best estimates of hydrologic state, particularly SWE.

The experiments presented here are motivated by the need to reduce state estimation errors in models used to initialize hydrological forecasts in an automated and reproducible manner. With that goal in mind, the results are generally encouraging, particularly for soil moisture states, but also highlight areas of concern. We have shown that assimilation of streamflow can improve simulated soil moisture and SWE state, depending on the basin characteristics, including concentration times and hydroclimate. In basins where streamflow offers little information to constrain SWE, assimilation of streamflow observations alone may not be able to reduce the uncertainty in mid-winter to early spring SWE states, and may even degrade those systems. While this may not have a large impact on short- to medium-range forecasts in such systems, it could reduce the skill of seasonal water supply forecasts where snow becomes a critical predictor for future runoff.

While the sequential particle filter generally outperformed the non-sequential filters proposed here, the non-sequential filters outperformed the open loop case, in terms of CRPSS, in most years for streamflow, soil moisture and SWE (except for SWE at Howard Hanson). The value of the non-sequential filters is enhanced if the streamflow observations have lower errors. The primary benefit to employing a non-sequential filter is that particle weights in a non-sequential filter are easier to interpret because the user specifies which simulated values influence their calculation to what degree. The user can control and limit the degree to which simulation performance memory is accounted for in evaluating particle performance.

We have also found that the use of a particle filter that is conditioned only on streamflow can degrade the quality of SWE states produced by a well-calibrated model that is run in open loop mode. This is particularly true if streamflow is driven by rainfall rather than snowmelt, as is the case at Howard Hanson.

This study did not attempt to calibrate model parameters through the DA process or estimate model structure and parameter uncertainty. If included, this added uncertainty would drive the particle filter to favor the observations more heavily. Because the non-sequential particle filter performed similar to the sequential filter when the observations were favored more heavily in the current simulation ($\sigma_t = 0.10z_t$), it is likely that incorporating model uncertainty would have led to a smaller discrepancy in CRPSS between the sequential and non-sequential filters; however, additional work is needed to examine this question.

While we have shown improvements in distributions of streamflow and soil moisture states with the sequential particle filter, SWE states were more variable. Future work will examine the degree to which the filter's changes to hydrologic moisture state distributions propagates into improved short- to medium-range streamflow forecasts in a real-world system. The variation of CRPSS for SWE with climate in the two basins studied here suggests that additional work should focus on a variety of hydroclimatological systems to provide greater insight on an aspect that remains particularly challenging. Based on the results presented here, the sequential particle filter provides a reasonable and automated approach to improve hydrologic states for use in the initialization of short- to medium-range streamflow forecasts, particularly in basins whose streamflow does not depend on snow. In basins where SWE is important, we recommend incorporating SWE observations as an additional constraint in the DA system, particularly when these states are also used in seasonal water supply forecasts.

Chapter 3. IMPLICATIONS OF STREAMFLOW DATA ASSIMILATION VIA A PARTICLE FILTER FOR STREAMFLOW FORECASTS IN WATERSHEDS WITH SEASONAL SNOWPACK

This chapter will be submitted to *Hydrology and Earth System Sciences* as

Clark, E. A., B. Nijssen, A. W. Wood, and M. P. Clark. Implications of streamflow data assimilation via particle filter on streamflow forecasts in basins with seasonal snow. *Hydrology and Earth System Sciences*, in preparation.

Abstract

Data assimilation (DA) has been widely studied as a key strategy to improve estimates of hydrologic states for use in hydrologic forecasting. Relatively few studies, however, have examined the performance of streamflow forecasts initialized with snow and soil moisture states that were updated only by the assimilation of streamflow observations. We employ the Sequential Importance Resampling Particle Filter (SIR-PF)—a DA method that can be used in non-linear and non-Gaussian systems without directly adjusting model states—to assimilate streamflow observations to improve probabilistic estimates (ensembles) of hydrologic model states. We then assess the accuracy and reliability of streamflow forecasts initialized with the resultant state distributions in two watersheds—the Green River Basin upstream of Howard Hanson Dam and the South Fork Flathead River upstream of Hungry Horse Dam—using 10 years of daily 7-day lead ensemble forecasts. In both basins, we find that the SIR-PF DA improves forecast skill (measured via several metrics relative to a baseline open loop state initialization) when using as few as five IHC ensemble members, when only the members with

the largest PF weights are selected for forecast initialization. Forecasts initialized from the SIR-PF with 25% uncertainty in its likelihood function outperform both open loop simulations and also the SIR-PF with 10% uncertainty, based on a range of metrics. Improvements due to DA decrease as a function of lead time, however, and at least one skill metric (correlation) is not improved through DA. The skill improvements are largest in spring and summer, which is consistent with improved SWE estimates during spring, when snow melt contributes most dynamically to streamflow. We also find significant year-to-year variations in DA performance, indicating that DA studies using shorter hindcast records may yield unreliable results.

3.1 INTRODUCTION

Short- to medium-range streamflow forecasts are a central information product that water managers use to determine how much water to store or release from reservoirs on a given day. The value of these forecasts decays rapidly as forecast latency increases. Therefore, it is important that streamflow forecasts be issued as close to real-time as possible. Streamflow forecasts (0-6 days) are primarily dependent on two factors: 1) how well we can predict the future meteorological conditions, and 2) how well we can estimate current watershed moisture storage conditions (primarily soil and snow). Measurements of soil moisture and snow water storage are sparse in both spatial density and temporal coverage. As a result, streamflow forecasters rely on hydrologic models forced by observed meteorology (e.g., air temperature, precipitation, and in some cases other model inputs) to simulate a watershed's initial hydrologic conditions (IHCs) at the start of the forecast. These simulated IHCs, like all hydrologic model outputs, are subject to several sources of uncertainty, including errors in past forcings (e.g., precipitation and temperature), model structural errors, and numerical errors [e.g., *Wagener and*

Gupta, 2005]. The existence of these errors is well known, but their magnitude, which varies in time and space, is not.

Data assimilation (DA) seeks to reduce simulation errors by incorporating additional watershed observations into model state estimates. Several methods of DA have been developed and tested in the field of hydrology, where the most popular methods of DA for prediction [*McLaughlin*, 2002; *Liu et al*., 2012] are variations on 1) the Ensemble Kalman Filter (EnKF) [*Evensen*, 2007], 2) the Particle Filter (PF) [*Kitagawa et al*., 1996], and 3) variational methods [*Daley*, 1991]. PFs are popular in hydrology because they do not require a Gaussian error distribution, they work well in non-linear systems [*Arulampalam et al*., 2002], and they do not modify the model states directly, thereby maintaining an internally consistent water balance [*Salamon and Feyen*, 2010]. A primary drawback of PFs is sample degeneracy, in which the majority of the weight is given to a very small subset of particles, such that the posterior distribution cannot be represented [*Li et al.*, 2014]. To address this issue, Sequential Importance Resampling (SIR) PF methods resample the particles when sample degeneracy occurs [*Doucet et al.*, 2001], but the side-effect of resampling is sample impoverishment, in which the ensemble spread is reduced for several time steps following the resampling [*Li et al*., 2014]. In hydrologic applications, the filter tends to collapse during periods of low variability in watershed conditions (e.g., during low-flow periods). Several additional methods have been developed to account for sample degeneracy and impoverishment, including residual resampling [*Zhang et al.,* 2013], implicit equal-weights [*Zhu et al.,* 2016], particle Markov chain Monte Carlo methods [*Andrieu et al.*, 2010; *Moradkhani et al.,* 2012], use of genetic algorithm operators [*Yin and Zhu*, 2015], hybrid PF-EnKF methods [*Slivinski et al*., 2015], and others. Despite these many methodological developments, PF methods remain relatively unused in hydrologic forecasting practice.

Forecasters remain concerned with the (perceived) computational burden of ensemble methods and remain unconvinced of the practical utility of PF methods. In this paper, we explore the utility of the simplest of these methods (SIR-PF) [*Arulampalam et al*., 2002; *Moradkhani et al*., 2005; *Weerts and El Serafy*, 2006] for improving short-range streamflow forecasts in watersheds with seasonal snowpack.

Several studies have demonstrated the use of PF to assimilate streamflow [*Dumedah and Coulibaly*, 2013; *Moradkhani et al*., 2005, 2006; *Weerts and El Serafy*, 2006], soil moisture [*Montzka et al.*, 2011] or both [*Yan and Moradkhani*, 2016] for hydrologic state and/or model parameter estimation. These studies primarily test the method in basins that have little to no snow. Other studies have used PF to assimilate observations of snow water equivalent (SWE) [*Leisenring and Moradkhani*, 2010; *DeChant and Moradkhani*, 2011a], snow depth [*Magnusson et al.*, 2017], and brightness temperature [*DeChant and Moradkhani*, 2011b; *DeChant and Moradkhani*, 2014] to improve hydrologic states in basins with seasonal snow cover. Still, few studies have assessed the benefits of assimilating only streamflow—the most readily available and accurately observed hydrologic variable in common observing systems—in basins where seasonal snow plays a strong role in the hydrologic cycle [*Abaza et al*., 2015]. In addition, these studies examine relatively short periods—the longest analysis periods are four [*Dumedah and Coulibaly*, 2013] and six years [*DeChant and Moradkhani*, 2014]. The former study focuses on parameter estimation and aggregate forecast performance, and the latter is a seasonal flow volume forecast experiment. As a result, neither discusses interannual variability in forecast performance.

Among the few studies that have addressed the topic of the assimilation of only streamflow in basins with seasonal snow cover, *Abaza et al.* [2015] used an EnKF approach

based on streamflow observations in two basins with strong snowmelt signals in Quebec, Canada. They found that updating both soil moisture and snow states in the Hydrotel hydrologic model produced better streamflow forecast accuracy and reliability—at lead times from 0 to 240 hours (10 days)—than updating either soil moisture or snow states separately. A forecast is reliable if the forecast and observations are statistically consistent (i.e., if the observed time series could have been drawn from the forecast distributions). The EnKF produced the greatest improvements relative to the open loop at short lead times (up to 2 days). Finally, *Abaza et al.* [2015] reported that a 50-member ensemble (created by randomly pairing 1 meteorological ensemble forecast member with each of 50 IHC ensemble members) could produce similar forecast performance as a 1000-member ensemble (with each of 20 meteorological forecast ensemble members paired with each of 50 IHC ensemble member).

In Chapter 2 of this dissertation, we applied multiple PF realizations in a synthetic truth experiment that examined the effects of assimilating only streamflow on simulated streamflow, soil moisture, and SWE model states. Because the PF does not adjust states directly, but rather alters their relative weighting at each update step, and resamples which ensemble members (particles) continue to evolve, soil moisture and SWE state values for a given particle remain consistent with model dynamics and forcings. We found that while soil moisture and streamflow were almost always improved by the assimilation of streamflow, SWE states—which are directly linked to streamflow during spring melt—can be degraded by the assimilation of streamflow, particularly during the middle of winter when little snowmelt occurs. Like *Abaza et al*. [2015], we noted a difference in performance between basins: the basin in which snow was the dominant form of precipitation (Hungry Horse) yielded more improvement in SWE than the basin in which precipitation fell as a mixture of rain and snow (Howard Hanson).

This paper addresses two primary questions. First, does PF DA of streamflow observations improve short-term streamflow forecasts—in terms of accuracy and reliability—relative to an open loop (no DA) case in basins with seasonal snow cover? And second, can we enhance the operational usability of PF DA by limiting the streamflow forecast ensemble size?

To answer these questions, we perform a series of hindcast experiments in the same two basins in the Pacific Northwest of the U. S. as used in Chapter 2. One of the basins (Howard Hanson) has precipitation that is a mixture of rain and snow, while the other (Hungry Horse) primarily receives winter precipitation as snowfall. We perform two hindcasting experiments using the SIR-PF, with assumptions of 10% and 25% uncertainty (which determines the likelihood of each simulated ensemble member), and one experiment without DA (the open loop). Each experiment generates a time-varying ensemble of IHCs. Higher uncertainty in the likelihood function allows the PF to retain a wider spread of ensemble members and implies that the model simulation is given relatively more weight with respect to the observations. We pair the resultant IHCs with an ensemble of meteorological forecasts, whose spread represents the uncertainty in atmospheric conditions within numerical weather predictions (NWP). Daily streamflow hindcasts are made four times per month for lead times of 0 to 6 days for nine water years from 2006-2015. We first examine the change in streamflow hindcast performance when only a subset of IHC ensemble members is used to initialize the forecasts. The results are compared in terms of three deterministic measures based on the ensemble mean—mean absolute error, percent bias, and correlation—and two probabilistic measures – Continuous Rank Probability Score (CRPS) [*Hersbach et al*., 2000] and α reliability [*Renard et al.,* 2010]. These measures are also compared both year-round at increasing lead times and seasonally. It is important to consider seasonal metrics because of large differences in hydrodynamics during

different regimes such as fall and winter snow accumulation, fall rainstorms, spring snowmelt, and summer low flows.

These methods are described in detail in section 3.2. Section 3.3 presents the results, including example forecast hydrographs, metrics as a function of the number of IHC ensemble members, the dependency of forecast performance on lead time and hydroclimate, and seasonal analyses. We discuss the implication of these results in section 3.4 and present concluding remarks and recommendations in section 3.5.

3.2 METHODS

3.2.1 *System design*

A broad overview of the hindcast system design employed in this study (Fig. 3.1) is described here. Details about each component are presented in the sections that follow. Hindcasting is a tool to evaluate forecast performance by reconstructing forecasts for a historic period. Hindcasts are historical forecasts, generated using the same forecast system and techniques from the real-time system to be evaluated, including the same data sources, model parameters and code. Hindcasting addresses two issues in forecast evaluation: 1) verification metrics require a long record of forecasts; and 2) future observations are not available to evaluate current forecasts. The results presented in this paper are based on hindcasts, but because the methodology mirrors that of a real-time forecast system, we use the terms interchangeably.

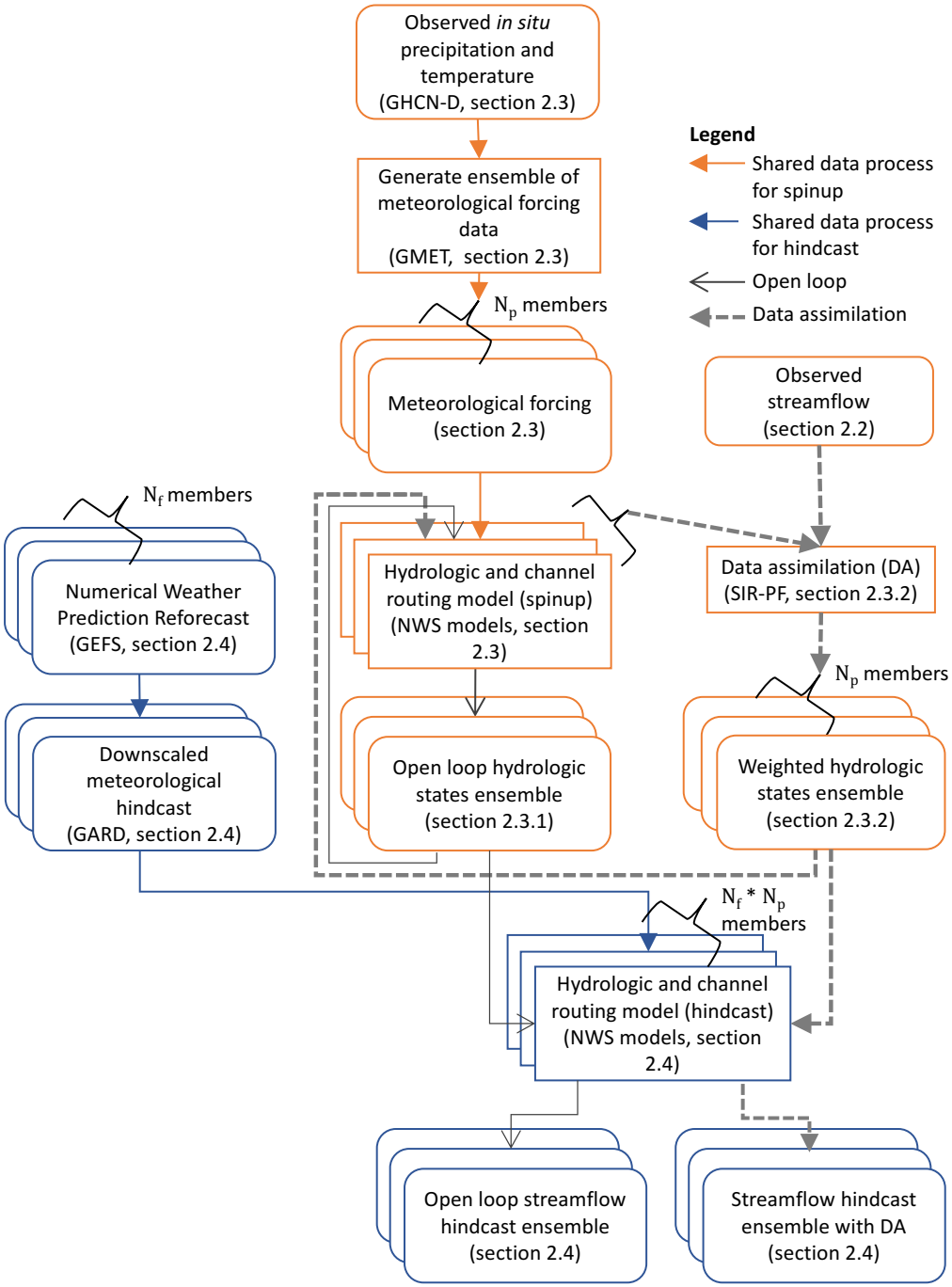

Figure 3.1. Workflow for generating hindcasts. Initial conditions are taken either from open loop hydrologic states ensemble or hydrologic states ensemble updated with DA. Meteorological forecasts are downscaled and used to force hydrologic models.

Streamflow forecasts are generated by a hydrologic model. Forecasts are initialized with an ensemble of simulated IHCs developed from an ensemble of meteorological forcings (section 3.2.3), and then forced in the forecast period by an ensemble of downscaled meteorological reforecasts (section 3.2.4). For each forecast, we pair each meteorological reforecast with each IHC, so that the number of streamflow forecast simulations performed on each day is equal to the product of the number of meteorological forecast ensemble members and the number of IHC ensemble members. Both the base case, or open loop, and the DA cases rely on the ensemble of IHCs. In the open loop case, the IHCs are considered equiprobable (section 3.2.3.1), whereas in the DA case, the IHC ensemble members are assigned weights representative of their probability given a streamflow observation (section 3.2.3.2).

3.2.2 *Study sites*

The western U.S. depends on water storage in reservoirs and snowpack for irrigation, navigation, municipal and industrial uses, flood control, environmental flows and hydropower. Hungry Horse Reservoir and Howard Hanson Dam are located in mountain headwaters in the Pacific Northwest, U.S. (Fig. 3.2). Hungry Horse is fed by the North Fork Flathead River, a tributary to the Snake River. The North Fork Flathead River drains an area of 4200 km^2 upstream of Hungry Horse, of which 83% is covered by forest [*McCarthy et al.,* 2016]. One third of the 1070 mm/year of precipitation at Hungry Horse falls between December and February (33%), and most of the 785 mm/year of runoff occurs during the spring (43%, March-May) and summer (42%, June-August). Howard Hanson is fed by the Green River, which is a tributary to the Duwamish River in the western Cascade Mountains. The Green River drains an area of 570 km^2 upstream of Howard Hanson, of which 91% is forested [*Sumioka et al*., 1998]. The Green River upstream of Howard Hanson receives water from both snowmelt and rainfall runoff. Fall

precipitation (September-November) accounts for 29% of the 1970 mm/year of precipitation, while another 37% occurs in December to February. Most of 1680 mm/year of runoff enters Howard Hanson in spring (38%) and winter (34%). This climatology is based on water years (October through September) 2007-2015 from the ensemble mean of precipitation observations described in section 3.2.3 and from streamflow observations. The primary differences between these basins are drainage area, and therefore time of concentration, and the main form of precipitation.

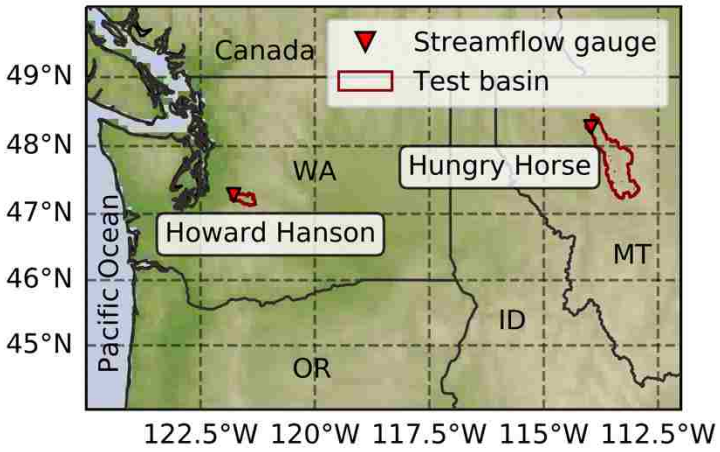

Figure 3.2. Location map of study basins.

Streamflow observations for assimilation and forecast evaluation were obtained from the U.S. Bureau of Reclamation AgriMet network for inflows to Hungry Horse (https://www.usbr.gov/pn-bin/arcread.pl?station=HGH) and from USACE Columbia Basin Water Management Division for inflows to Howard Hanson (http://www.nwd-wc.usace.army.mil/dd/common/dataquery/www). Because the study basins are in the headwaters, both experience minimal human intervention (e.g., irrigation, groundwater pumping).

3.2.3 *Watershed initial hydrologic conditions (IHCs) for forecasting*

Daily timestep hydrologic states are simulated using the Sacramento Soil Moisture Accounting model [*Burnash et al*., 1973] and the Snow-17 model for snow accumulation and ablation [*Anderson*, 1973], while drainage and melt are converted to streamflow using a unit hydrograph routing model. Because these three models are run in combination by the U.S. National Weather Service (NWS) River Forecast Centers (RFCs) in generating operational streamflow forecasts, we refer to the combination as the NWS models. They were selected specifically to evaluate the utility and feasibility of ensemble methods in operational settings. Each study basin is represented in the NWS models by two hydrologic response units based on elevation. The NWS models are run from 1970-2006 to "spin up" the model's storage components; the particle filter is run for water year 2006 to allow the weights to adjust; and the hindcast analysis period spans water years 2007-2015.

To represent uncertainty in IHCs, 100 NWS model simulations are forced with each of 100 unique and equally probable gridded meteorological forcing ensemble members (Chapter 2). The forcings are derived using a real-time version of the Gridded Meteorological Analysis Tool (GMET) [*Newman et al.,* 2015]. GMET performs a probabilistic interpolation of in situ observations of daily precipitation and temperature, based on spatial regression against location, elevation, aspect, latitude and longitude [*Clark and Slater*, 2006; *Newman et al.,* 2015].

The ensemble state distribution is representative of the errors in both in situ observations of precipitation and temperature and in spatial interpolation. In situ observations of daily precipitation and minimum and maximum temperature are extracted from the Global Historical Climatology Network-Daily dataset [*Menne et al*., 2012a, 2012b]. Station data are quality controlled, gap-filled, and used to construct daily ensemble gridded 1/16° forcings from 1/1/1970

to present. Then, they are mapped through conservative regridding to the NWS model elevation zones.

We construct a single forcing time series for calibration of the NWS models by randomly selecting year-long segments of different meteorological forcing ensemble members. As noted in in Chapter 2, this process reduces model bias when the model is run with the entire forcing ensemble. We calibrate model parameters in each basin over the period of 1981-2008 to naturalized streamflow observations from the No-Regulation No-Irrigation (NRNI) streamflow dataset developed by the Columbia River Management Joint Operating Committee [*CRMJOC*, 2015]. We use the Multi-Objective Complex evolution Optimization Model (MOCOM) [*Yapo et al*., 1998] with a dual objective function that minimizes the maximum long-term mean monthly streamflow error and the root mean square error (RMSE) of daily streamflow. Nash-Sutcliffe Efficiency scores for daily streamflow values during the calibration period are acceptable but not outstanding—0.70 and 0.86 for the Howard Hanson and Hungry Horse basin models, respectively.

3.2.3.1 Open loop

The open loop (or no DA) simulation consists of 100 unique daily simulations of hydrologic states and fluxes. Each ensemble member in the open loop case uses a single meteorological forcing ensemble member for all days during both the spin-up and analysis periods. We subset this ensemble when selecting IHCs for forecasting based on the ensemble median simulated streamflow (section 3.2.4).

3.2.3.2 Data assimilation

In this study, we use the SIR-PF implementation described in Chapter 2 to assimilate streamflow into NWS model simulations. This implementation is adapted from *Arulampalam et al.* [2002],

Moradkhani et al. [2005], and *Weerts and El Serafy* [2006]. A brief overview of the technique is presented here. In SIR-PF, an ensemble of model simulations is run, as in the open loop case, to represent the system's uncertainty. Each member is referred to as a "particle", which is a weighted ensemble member. We use these terms interchangeably. The weights represent the relative contribution of each particle to the probability distribution of the ensemble, which is dependent on its agreement with observations. *Kitagawa* [1996], *Doucet et al.* [2001], and *Arulampalam et al.* [2002] show that the posterior probability function ($p_t(\mathrm{x})$ at time t) for the modeled ensemble values (x_t), given the observations ($z_{1:t}$) can be approximated as:

$$p_t(x_t|z_{1:t}) \approx \sum_{i=1}^{N_p} w_t^i \, \delta\left(x_t - x_t^i\right), \tag{3.1}$$

where x_t^i and w_t^i are the states and normalized weights, respectively, at time t for particle i, for a total of N_p particles, and $\delta(\cdot)$ is the Dirac delta function. In Eqn. 3.1, x_t is the vector of random variables that forms the x-axis of the probability distribution. The delta function equals one for x_t equals x_t^i and zero for all other values. The resulting cumulative distribution function looks like a staircase with an increase in probability of height w_t^i at each $x_t = x_t^i$.

New observations are compared to each ensemble member in turn and the likelihood of the observation given the simulated value is estimated. In this study, as in *Moradkhani et al.* [2005] and Chapter 2, the likelihood ($p_t\left(z_t|x_t^i\right)$) of the observed streamflow (z_t) at time t, given the modeled streamflow for particle i, is calculated as a Gaussian function of the difference between the observed value and the modelled value, with a standard deviation of σ_t:

$$p\left(z_t|x_t^i\right) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{(z_t - x_t^i)^2}{2\sigma_t^2}}. \tag{3.2}$$

The standard deviation represents an estimate of the combined model and observational uncertainty and is varied as described in section 3.2.4. Next, in SIR-PF, the particle weights for time t are calculated as function of the likelihood Eqn. 3.2) and the previous time step's weight [*Doucet et al.,* 2001]:

$$w_t^i \propto w_{t-1}^i \cdot p\left(z_t \middle| x_t^i\right). \tag{3.3}$$

Combining Eqns. 3.2 and 3.3, the new weight for each particle at each timestep, w_t^{i*}, is

$$w_t^{i*} = w_{t-1}^i \cdot \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{\left(z_t - x_t^i\right)^2}{2\sigma_t^2}} \cdot \left(\sum_{m=1}^{N_p} \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{\left(z_t - x_t^m\right)^2}{2\sigma_t^2}} \right)^{-1} \tag{3.4}$$

which is then re-normalized so that all the ensemble weights sum to one:

$$w_t^i = \frac{w_t^{i*}}{\sum_{m=1}^{N_p} w_t^{m*}}. \tag{3.5}$$

If the weights are highly disparate, such that all weight is given to a number of particles too small to represent the distribution (an effect known as sample degeneracy) [e.g., *Li et al.,* 2014], particles are resampled. We calculate the number of effective particles (N_{eff}, the number of particles that contribute meaningfully to the posterior distribution) as described in *Doucet et al.* [2001]:

$$N_{eff} = \frac{1}{\sum_{i=1}^{N_p} {w_t^i}^2}. \tag{3.6}$$

In preliminary tests, we noted that frequent resampling produced overconfident ensembles (i.e., the ensemble spread was smaller than indicated by observations). This was particularly problematic in dry or cold periods when the streamflow was not variable and the observed streamflow was biased for prolonged periods. To avoid this, we choose to resample whenever N_{eff} falls below $N_p/5$ because it leads to less frequent resampling compared to resampling at every time step as in *Moradkhani et al.* [2005]. Alternative thresholds may be

appropriate as well. We follow the resampling method outlined in Chapter 2, which is based on the algorithm presented in *Arulampalam et al.* [2002], *Moradkhani et al.* [2005], and *Weerts and El Serafy* [2006]. This method reconstructs the posterior distribution of streamflow by resampling from the effective subset of particles with a frequency matching their weights. Particles with very low weights may be discarded during resampling, while the states of the highest weighted particles may be assigned to more than one particle in the resampled distributions. When more than one particle is assigned the same state during resampling, the new particles are assigned forcings from particles that were dropped during resampling.

3.2.4 *Hindcasts*

The hindcast system used in this study (Fig. 3.1) has been implemented as part of a System for Hydromet Analysis, Research and Prediction (SHARP) [*Wood et al.*, 2016], leveraging several components that support experimentation in both historic and real-time ensemble prediction [for more details, see Chapter 2 and *Mendoza et al*., 2017]. Ensemble streamflow hindcasts are initialized with IHCs from section 3.2.3 and forced with an ensemble of meteorological reforecasts of precipitation and temperature from NOAA's National Centers for Environmental Prediction (NCEP)'s Global Ensemble Forecast System (GEFS) [*Buizza et al.,* 2005] Reforecast-2 [*Hamill et al.*, 2013]. As with the historic forcings, each of the 11 forecast ensemble members as well as the mean were regridded to a 1/16° spatial resolution using the Generalized Analog Regression Downscaling tool (GARD) [*Gutmann et al.*, 2017] and conservatively mapped to polygons for two hydrologic response units for each basin. Hindcasts are generated for the 1st, 8th, 16th, and 24st day of each month from water year 2006 to 2015, resulting in 432 total hindcasts with forecast lead times from 0-6 days.

Combining all 100 IHC ensemble members with all 11 GEFS ensemble members results in a total of 1100 streamflow forecast ensemble members. Ensemble size is a common constraint for operational real-time systems, so we test the performance of streamflow forecasts in which each of 11 meteorological forecasts is initialized with smaller ensembles of N_{IHC} members, where N_{IHC} varies from 1 to 100. In the open-loop case, we use the N_{IHC} median IHC ensemble members (e.g., if N_{IHC}=15, the ensemble members with rank 43-57 (out of 100) are used). Because ensemble distributions can be skewed, the median members do not necessarily have the flows that are closest to the ensemble median flow. In the DA cases, we use the N_{IHC} IHC ensemble members with the highest particle weights.

SIR-PF uses a likelihood function (Eqn. 3.2) to weight the IHC ensemble members. If the uncertainty σ_t in this equation is high, a larger range of particles will be given higher weights than if the uncertainty is low. This term is often defined as the uncertainty in the streamflow observations. For most gauges in the U.S., observed daily average streamflow is estimated to be accurate to within 2% to >8% of true streamflow. With our implementation of the SIR-PF, it is not possible to maintain an ensemble that represents the true uncertainty of the forecasting system. If too many particles are dropped because of the tight uncertainty bounds, then the filter cannot adapt if the model later misrepresents the system. For these reasons, model bias and uncertainty also impact the likelihood of an observation given the simulated values. Because these factors are difficult to quantify, some experimentation is required to identify an adequate level of uncertainty. Therefore, in addition to determining a reasonable minimum N_{IHC}, we also consider the impact of uncertainty on forecast performance by comparing hindcasts for which the combined observation and model uncertainty in the likelihood function (σ_t in Eqn. 3.2) is 10%

and 25%, respectively, of observed streamflow on day t. Higher uncertainty allows a larger subset of particles to be considered feasible as the ensemble evolves.

3.2.5 *Operational considerations*

Because the hindcasts are intended to test for performance implications of system changes (like the implementation of SIR-PF), it is important that they also reflect realistic operational settings. SHARP is an 'over-the-loop' forecasting system in which human forecasters do not directly adjust data or model outputs [*Wood et al.*, 2016]. It follows the workflow in Fig. 3.1 to produce ensemble streamflow forecasts in real-time. The amount of computing time increases with each additional ensemble member, as does file input and output (I/O). Adding an IHC ensemble member increases computing time in generating the meteorological forcing ensemble and running the hydrologic model for spinup and forecast. When the SIR-PF is included, additional IHC ensemble members also slows the calculation of weights and the resampling steps. The NWS model is a relatively quick hydrologic model in terms of computation time, which was a factor in its selection for this study; however, in expanding to a regional domain, a more distributed, higher dimensional model would likely be able to better represent hydrologic variability.

In the case of the NWS model, for each HRU, each additional IHC ensemble member adds one additional meteorological forcing file, one additional output file and three additional state files from the NWS model spinup. It also adds 11 (the number of meteorological forecast members) additional NWS output files from the hydrologic forecast. This is a total of 16 files per IHC ensemble member per HRU. In this paper, we consider only two basins with a total of four HRUs, but if we were to apply the NWS model across the U.S., increases in ensemble size would notably increase I/O. In this study, we used 100 IHC ensemble members because *Newman et al.*

[2015] demonstrated that an ensemble of this size represents the uncertainty in gridded forcing reasonably well.

3.2.6 *Evaluation metrics*

Forecast skill for this study is assessed in terms of three deterministic metrics: correlation (Pearson's R), mean absolute error (MAE), and percent bias (PBIAS), and two probabilistic metrics: continuous ranks probability score (CRPS) [*Hersbach*, 2000] and α-reliability index [*Renard et al*., 2010]. Each metric is calculated for all forecasts (4 forecasts per month for 9 years, totalling 432 forecast-observation pairs), each season (108 forecast-observation pairs per season), and each water year (48 forecast-observation pairs per year). All measures are also calculated as a function of the number of IHC ensemble members.

Because most existing systems that ingest ensemble traces treat each ensemble member as equal, and because the subsetting of IHC members already accounts for particle weights, we calculate validation statistics as if all streamflow forecast members are equally likely. Therefore, the deterministic metrics are calculated from the ensemble mean forecast for each day. MAE, which represents the volumetric errors in the mean streamflow forecast, is calculated as the average absolute difference between the ensemble mean forecast and the observed streamflow. For ease of comparison between basins, we also calculate the relative MAE (rMAE) by dividing MAE by the mean of observed streamflow for the same forecast-observation pairs. Pearson's R measures the degree of correlation in timing between observed streamflow and the ensemble mean forecast. Positive (negative) PBIAS shows an over- (under-) prediction of streamflow by the ensemble mean forecast.

Probabilistic metrics are calculated based on the distribution of streamflow forecast ensemble members. CRPS assesses the ability of the forecast probability distribution to capture

the observed flow. It compares the forecast cumulative distribution function ($F_t^s(s)$) to the distribution of observed flow ($F_t^y(s)$) for each forecast-observation pair [*Hersbach*, 2000]:

$$\text{CRPS} = \frac{1}{n}\sum_{t=1}^{n}\int_{s=-\infty}^{\infty}\left(F_t^s(s) - F_t^y(s)\right)^2 ds, \tag{3.7}$$

The observation in this case is assumed to be perfect, so the $F_t^y(s)$ is represented by the Heaviside function:

$$F_t^y(s) = \begin{cases} 0, s < y_t \\ 1, s \geq y_t \end{cases}. \tag{3.8}$$

CRPS rewards a close match between the forecast mean and the observation and penalizes ensemble spread. To enable comparison between basins and seasons, we convert CRPS to a Continuous Ranked Probability Skill Score (CRPSS):

$$\text{CRPSS} = 1 - \frac{\text{CRPS}}{\text{CRPS}_{ref}}, \tag{3.9}$$

where CRPS_{ref} is the CRPS of the corresponding open loop case. A perfect forecast has CRPSS of one. CRPSS is zero if there is no improvement from the open loop and negative if the forecast is worse.

Forecast reliability depends on statistical consistency of the time series of observed streamflow with the predicted probability distributions. Reliability is often examined by plotting predictive quantile-quantile (Q-Q) plots of the theoretical p-values derived from the uniform distribution ($p_{y_t}^{(th)}$) against the p-values of the observed streamflow (p_{y_t}) [*Laio and Tamea*, 2007]. p_{y_t} is derived from $F_t^s(s)$ for each forecast-observation pair. The α index was developed to quantify the reliability of forecast probability distributions [*Renard et al.*, 2010]. The index α measures the area between the forecast p-value curve and the 1:1 line. It varies between 0 (all observations outside of predicted range) and 1 (perfectly reliable) and is calculated as:

$$\alpha = 1 - 2\sum_{t=1}^{n} |p_{y_t} - p_{y_t}^{(th)}|/n. \quad (3.10)$$

Unlike CRPS, which rewards a small ensemble spread, α rewards an ensemble spread that is neither too narrow (overconfident) nor too wide (underconfident) than the total forecast uncertainty warrants.

3.3 RESULTS

When all 1100 streamflow forecast ensemble members are used, DA with 25% uncertainty improved forecasts—in terms of the total MAE over all seven days of lead time—relative to the open loop case for 60% of the forecasts at Hungry Horse and 55% of the forecasts at Howard Hanson. Example forecast hydrographs from the DA with 25% uncertainty and the open loop cases are shown for both study sites in Fig. 3.3. The upper panel shows the DA's best forecast, based on rMAE, averaged over all lead times. The middle panel shows the forecast at each basin that was most improved by DA, based on CRPS, averaged over all lead times. The lower panel shows the forecast at each basin that was most degraded by DA based on CRPS, averaged over all lead times. Fig. 3.3 shows that the influence of IHCs (i.e., distance between open loop and DA hindcasts) decreases as the lead time increases. In some cases, all ensemble members are biased relative to observations (e.g., Fig. 3.3c). In these cases, SIR-PF can minimize the ensemble mean bias, but the degree of bias correction is limited because states are not perturbed by this method. In other cases, all initialization methods miss a high flow event (e.g., Fig. 3.3d, lead-2) because the meteorological forecast missed the related precipitation event. Still, in most cases, the ensemble spread is reduced by DA (Fig. 3.3a, c, d, e), and in many cases, the bias of the ensemble mean is also reduced (Fig. 3.3a, b, c, d).

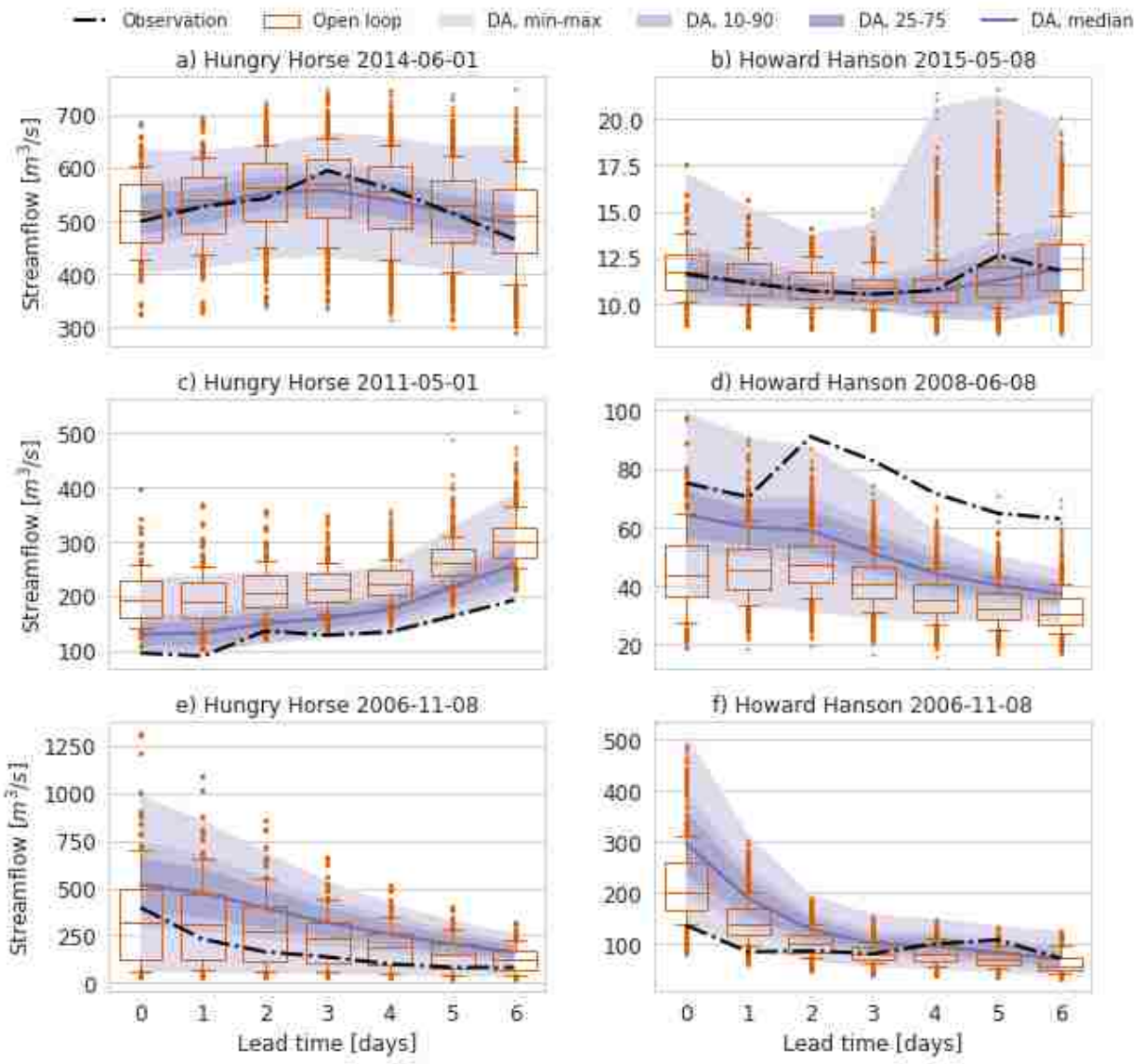

Figure 3.3. Ensemble streamflow forecasts for Hungry Horse (left) and Howard Hanson (right) showing the distribution of all 1100 streamflow forecast ensemble members for the open loop case and DA with 25% uncertainty in the likelihood function. a) and b) show the 7-day forecasts with the lowest relative MAE after DA; c) and d) show forecasts for which CRPS was most improved by DA; and e) and f) show the forecasts for which CRPS was most degraded by DA. For the open loop case boxplots (orange), the midline is the median, the box edges are 25th and 75th percentiles, whiskers show 10th and 90th percentiles, and the remaining ensemble members are plotted as dots. For DA, progressively lighter shading indicates 25th and 75th percentiles, 10th and 90th percentiles, and the envelop of minimum and maximum flow.

Over 3287 days in water years 2007-2015, at Hungry Horse the particles were resampled 704 times for DA with 10% uncertainty, but only 300 times when DA used 25% uncertainty. Similar frequencies of resampling occurred at Howard Hanson—784 times for 10% uncertainty and 298 times for 25% uncertainty. This means that the particle distribution degenerated roughly twice as often when we assumed 10% uncertainty (forcing the model to more closely reproduce observations) than when we assumed 25% uncertainty. Each time the particles are resampled, the ensemble spread decreases. More frequent resampling also requires more computer time, which we seek to limit in real-time forecasting systems.

3.3.1 *Sensitivity to the number of IHC ensemble members (IHC uncertainty)*

We compare CRPS for forecasts in which each of 11 meteorological forecasts was initialized with each of 1 to 100 IHC ensemble members (Fig. 3.4). At both study sites, in the open loop case, CRPS steadily decreases (improves) as more ensemble members are added. For DA with 10%, CRPS is worse than that of the open loop if more than 30 IHC ensemble members at Hungry Horse are used. The same is true at Howard Hanson if forecasts are initialized with more than 50 IHC ensemble members. At both basins, the CRPS of DA with 25% outperforms that of the open loop case for any subset of IHC ensemble members. In addition, at both basins for DA at each level of uncertainty, the improvement in CRPS with number of IHC members is minimal beyond approximately five IHC ensemble members. These findings suggest that, in this case, no more than five IHC ensemble members are useful in initializing the streamflow forecast. For this reason, the remaining results are presented at five IHC ensemble members (55 streamflow forecast ensemble members). Reasons for these findings are discussed in section 3.4.

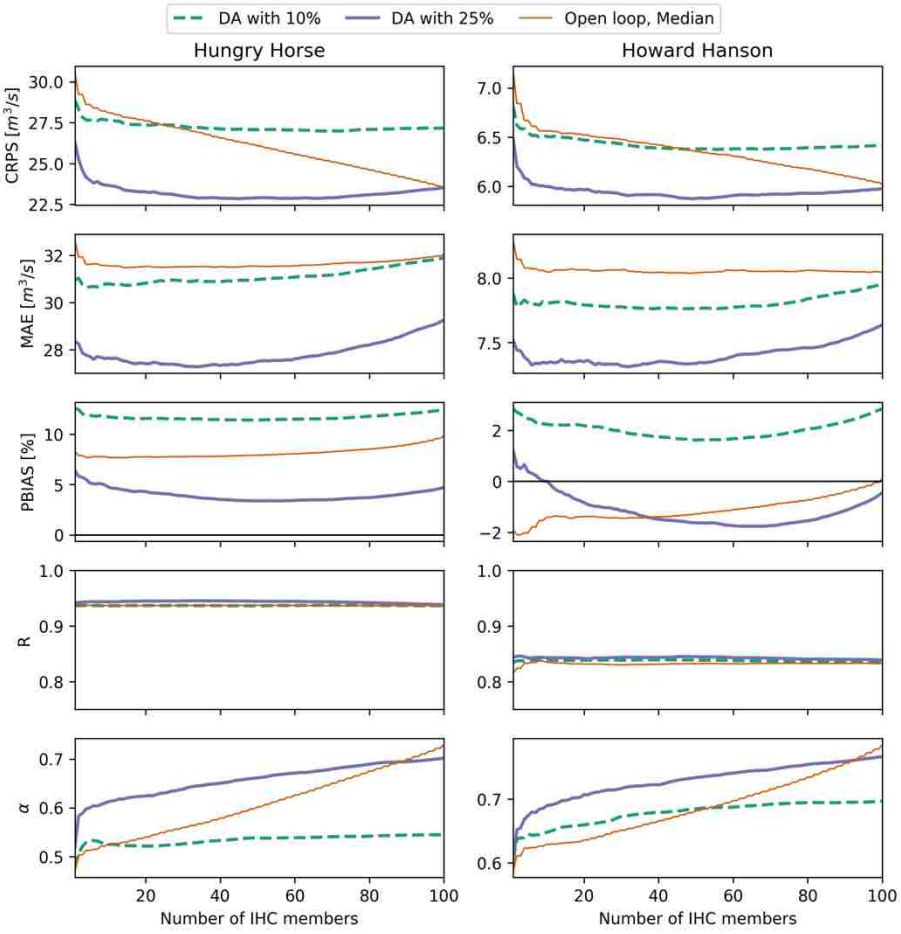

Figure 3.4. (Top to bottom) CRPS, MAE, PBIAS, Pearson's R, and α versus the number of IHC ensemble members used to initialize streamflow forecasts for Hungry Horse (left) and Howard Hanson (right).

3.3.2 *Sensitivity across lead-times, hydroclimate, and season*

Forecast verification metrics for both basins using five IHC ensemble members and 11 meteorological ensemble members are shown in Figs. 3.5 and 3.6 for DA with 10% and 25% uncertainty levels in the PF likelihood function (Eqn. 3.2), as well as for the open loop, compared with the scores resulting from a perfect forecast. Because forecast variations lead to both positive and negative variations across a suite of skill metrics, we introduce multi-attribute diagrams (in the form of radar plots) summarizing five metrics at once. The axes in Figs. 3.5 and

3.6 are oriented such that the best value of each metric, except for PBIAS, is on the outer circle, with worse values toward the center of the circle. The dashed line indicates a perfect forecast. For PBIAS, the perfect forecast (PBIAS=0%) is located at the midpoint of the axis. On the PBIAS axis, deviation towards the center of the diagram is negative, while deviation towards the outer edge of the diagram is positive. CRPSS is zero where the open loop line crosses its axis, and deviation from this point towards the center is negative (worse than the open loop).

Fig. 3.5 shows the forecast verification metrics calculated for all forecasts at lead times of 0, 2, 4, and 6 days. In both basins, all metrics worsen as lead time increases. Correlation is the least sensitive metric to the method of selecting IHC ensemble members. At Hungry Horse, correlation is very high (close to 1) for all methods and lead times. At Howard Hanson, correlation declines from approximately 0.8 at lead-0 to 0.5 at lead-6 for all methods. α varies more between methods than correlation, and is highest for DA with 25% at both basins for all lead times. PBIAS is negative at Howard Hanson positive at Hungry Horse for most lead times and initialization strategies. At Hungry Horse, PBIAS is best for DA with 25% at all lead times longer than 0 days; for lead-0, PBIAS is best for the open loop. At Howard Hanson, beyond lead-0, for which DA with 10% has the best PBIAS, all methods have nearly equal PBIAS. At Hungry Horse, rMAE is improved by DA with 10% and 25% at lead-0; beyond lead-0, DA with 25% has the lowest rMAE, and by lead-6, DA no longer improves rMAE relative to the open loop. At Howard Hanson, all methods produce the same rMAE at lead times longer than 0 days; at lead-0, the two DA cases outperform the open loop. For CRPS in both basins at lead-0, DA with 10% performs similarly to DA with 25%, and both outperform the open loop. At Hungry Horse, DA with 25% has the lowest CRPS at all remaining lead times, but at Howard Hanson, CRPS is nearly the same for all methods after lead-0.

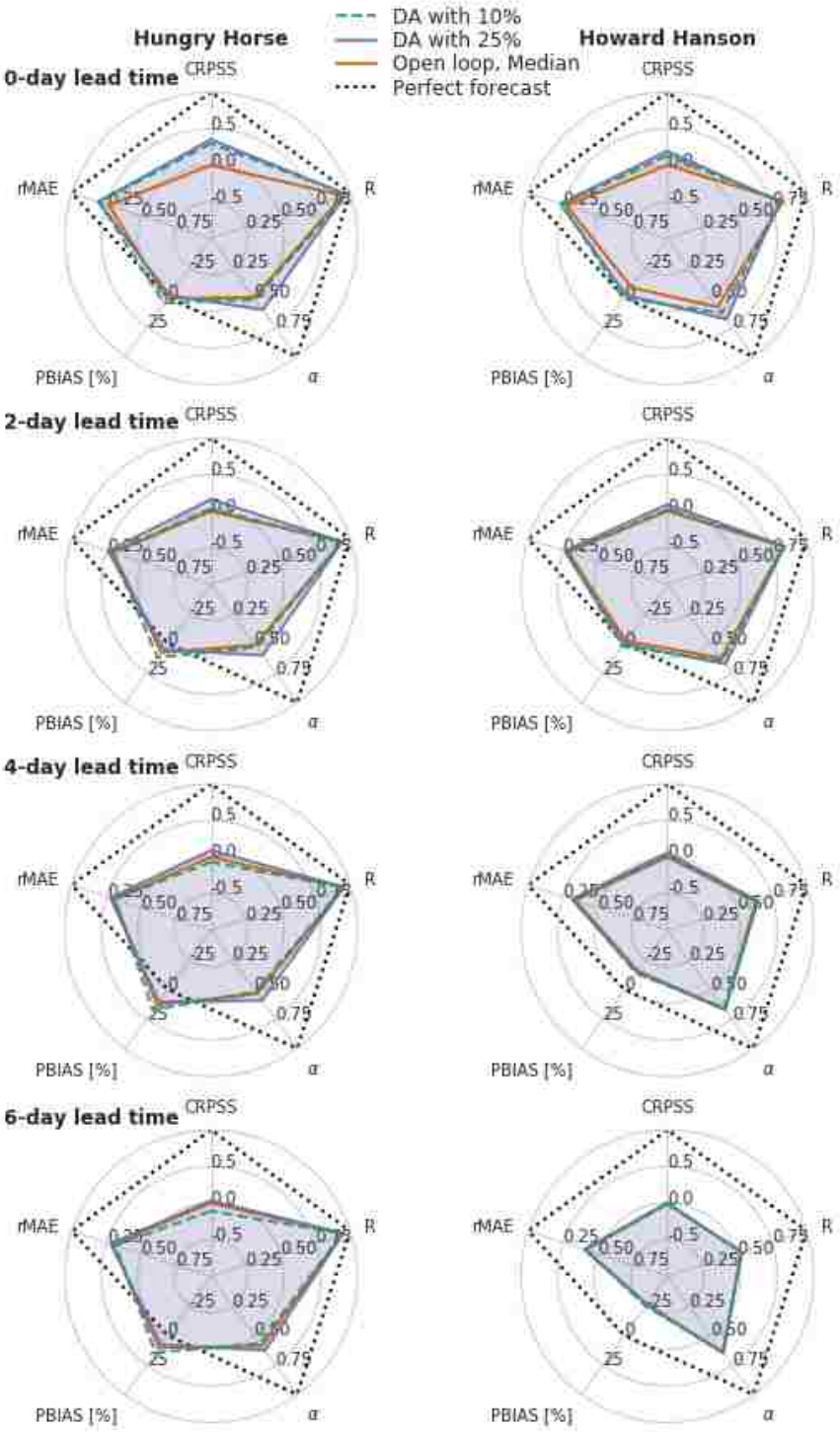

Figure 3.5. Multi-attribute verification plots showing CRPS, Pearson's R, α, PBIAS, and MAE for Hungry Horse (left) and Howard Hanson (right), calculated at lead times of 0, 2, 4, and 6 days ((top to bottom). All metrics calculated from five IHC ensemble members.

Metrics for each season are presented for five IHC ensemble members at a lead time of 2 days in Fig. 3.6. Fig. 3.6 shows that the largest overall improvement in forecast performance due to the inclusion of DA occurs in the spring for both basins. Correlation is the least sensitive metric to the method of IHC selection in all seasons. This occurs because the meteorological forecast is so highly correlated with flows in this basin that IHCs have little influence. The only notable difference in correlation between methods occurs in the winter at Hungry Horse. Correlation is close to one at both basins during the summer and at Hungry Horse during spring. At both basins in winter and fall, CRPS and rMAE are also insensitive to method of IHC selection. DA improves them most in spring and summer at Hungry Horse. DA with 25% has the highest α value in all cases except for summer at Howard Hanson, when α is the same for all 3 methods. In winter and spring, the highest α at Howard Hanson is higher than the highest α at Hungry Horse; while the highest α is comparable between basins in summer and fall. The open loop outperforms DA with 10% at Hungry Horse in the summer for all metrics. For both basins, the open loop outperforms DA for PBIAS in winter.

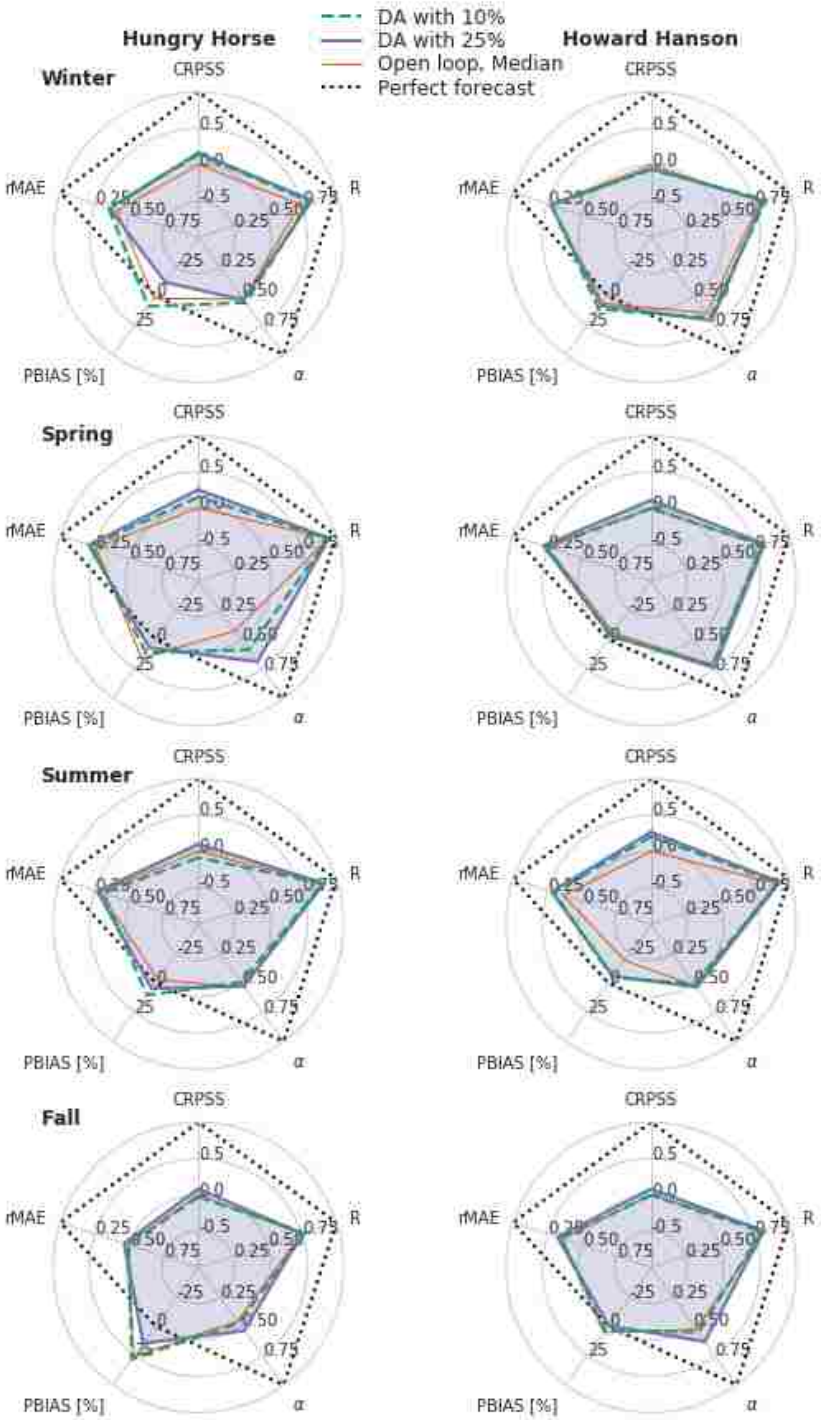

Figure 3.6. Multi-attribute verification plots showing CRPSS, Pearson's R, α, PBIAS, and rMAE for Hungry Horse (left) and Howard Hanson (right). Top to bottom: winter (DJF), spring (MAM), summer (JJA), and fall (SON). All metrics calculated at 2-day lead time initialized from five IHC ensemble members.

3.3.3 *Interannual variability*

During most years, DA with 25% uncertainty improves two or more forecast metrics relative to the open loop case at both basins (Figs. 3.7 and 3.8). Differences in polygon shape in Figs. 3.7 and 3.8 reflect the degree to which the performance of any given method varies from year to year.

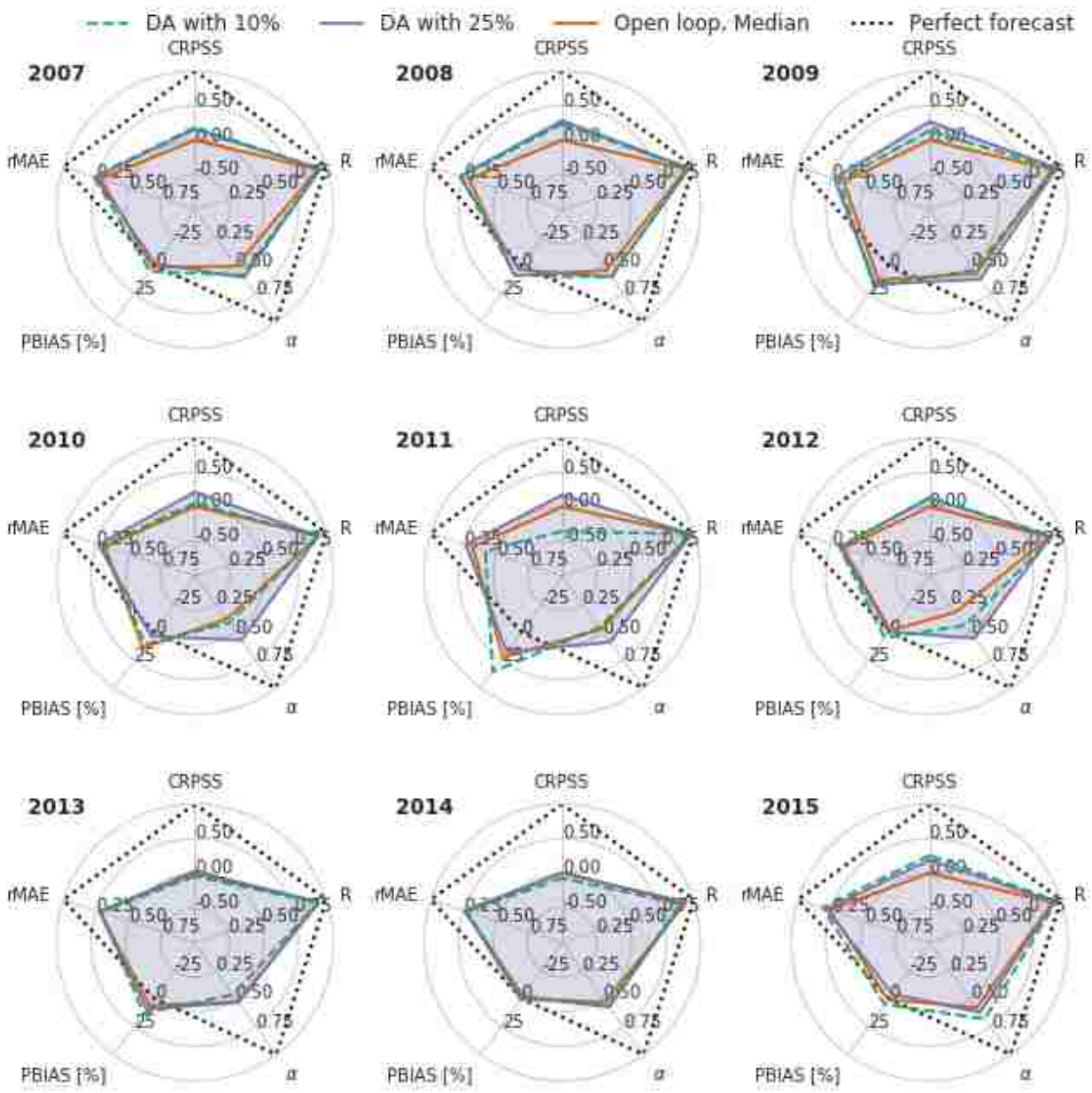

Figure 3.7. Multi-attribute verification plots showing CRPSS, Pearson's R, α, PBIAS, and rMAE for Hungry Horse for water years 2007-2015. All metrics calculated at 2-day lead time initialized from five IHC ensemble members.

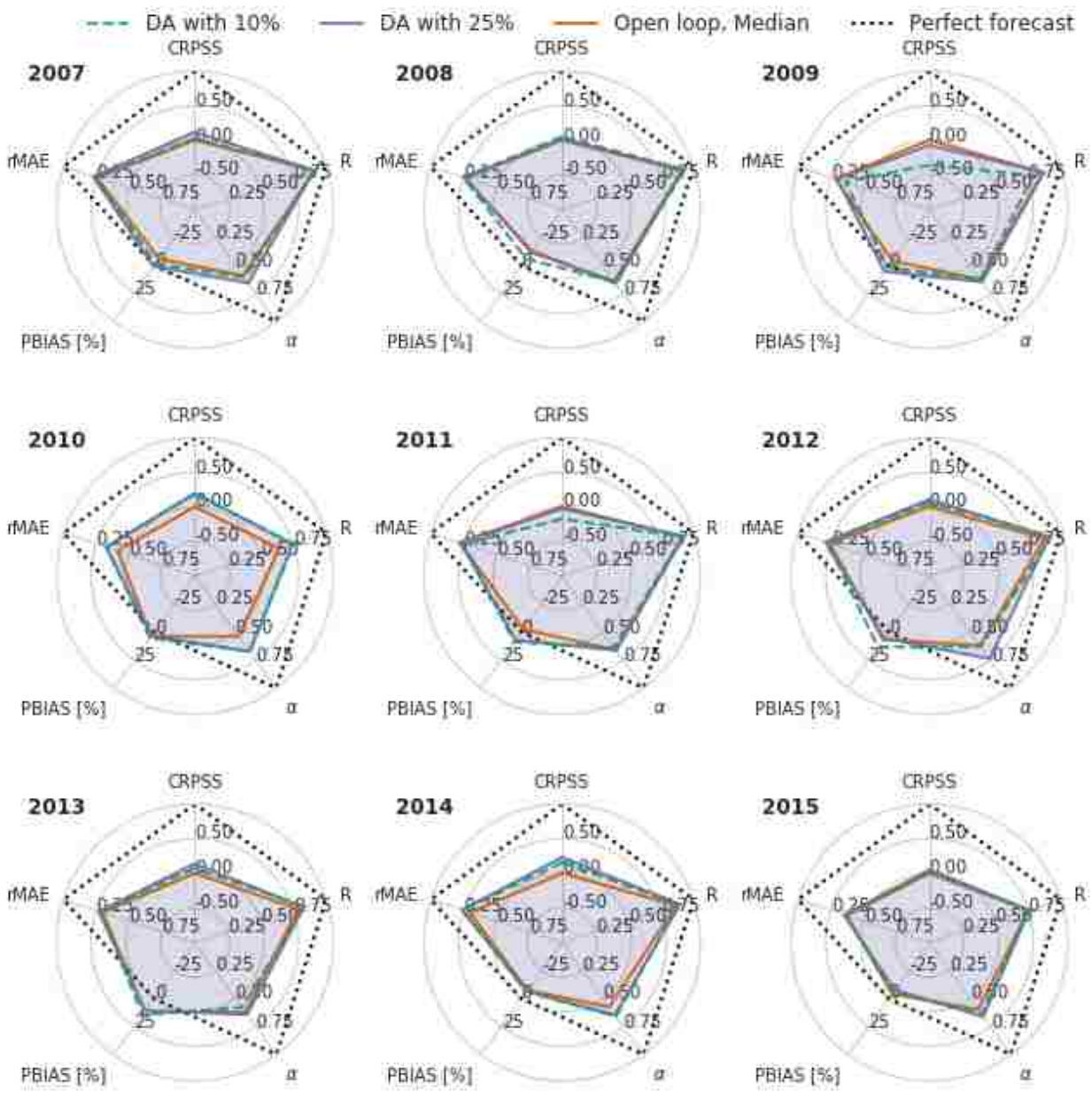

Figure 3.8. Multi-attribute verification plots showing CRPSS, Pearson's R, α, PBIAS, and rMAE for Howard Hanson for water years 2007-2015. All metrics calculated at 2-day lead time initialized from five IHC ensemble members.

For example, the shape of the open loop polygon at Howard Hanson is smaller for 2010 than for other years, reflecting worse performance in terms of rMAE, α, and R (Fig. 3.8). This is the only case in either basin that shows notable improvement in R due to DA, suggesting that

correlation improvements are only possible when the meteorological forecasts (the open loop case) are poorly correlated with observations. Howard Hanson's second lowest annual flow and precipitation occurs in water year 2010 (Fig. 3.9). Interestingly, the lowest flow at Howard Hanson occurs in 2015, and there is almost no improvement in forecast performance due to DA that year (Fig. 3.9). The later year is much warmer than 2010. El Niño conditions persisted throughout 2015, while there was a switch from El Niño to La Niña during 2010. It is likely that the meteorological forecasts were unable to adapt quickly to this shift in 2010, leading to the poor performance.

At Hungry Horse, DA with 10% uncertainty degrades at least two metrics (CRPSS and rMAE) relative to the open loop case in three of the nine years (Fig. 3.7); this only happens in 2011 at Howard Hanson (Fig. 3.8). 2011 is the wettest water year in the study period for Howard Hanson (Fig. 3.9). In some years (2015 at Hungry Horse, 2008 at Howard Hanson), DA with 10% uncertainty outperforms DA with 25% uncertainty in two or more metrics. At Hungry Horse, 2015—an El Niño period— was the driest water year during the study period and one of the warmest, while 2008—a La Niña period—was unremarkable at Howard Hanson (Fig. 3.9). These results suggest that the level of uncertainty necessary for good DA performance is not closely tied to climate variations.

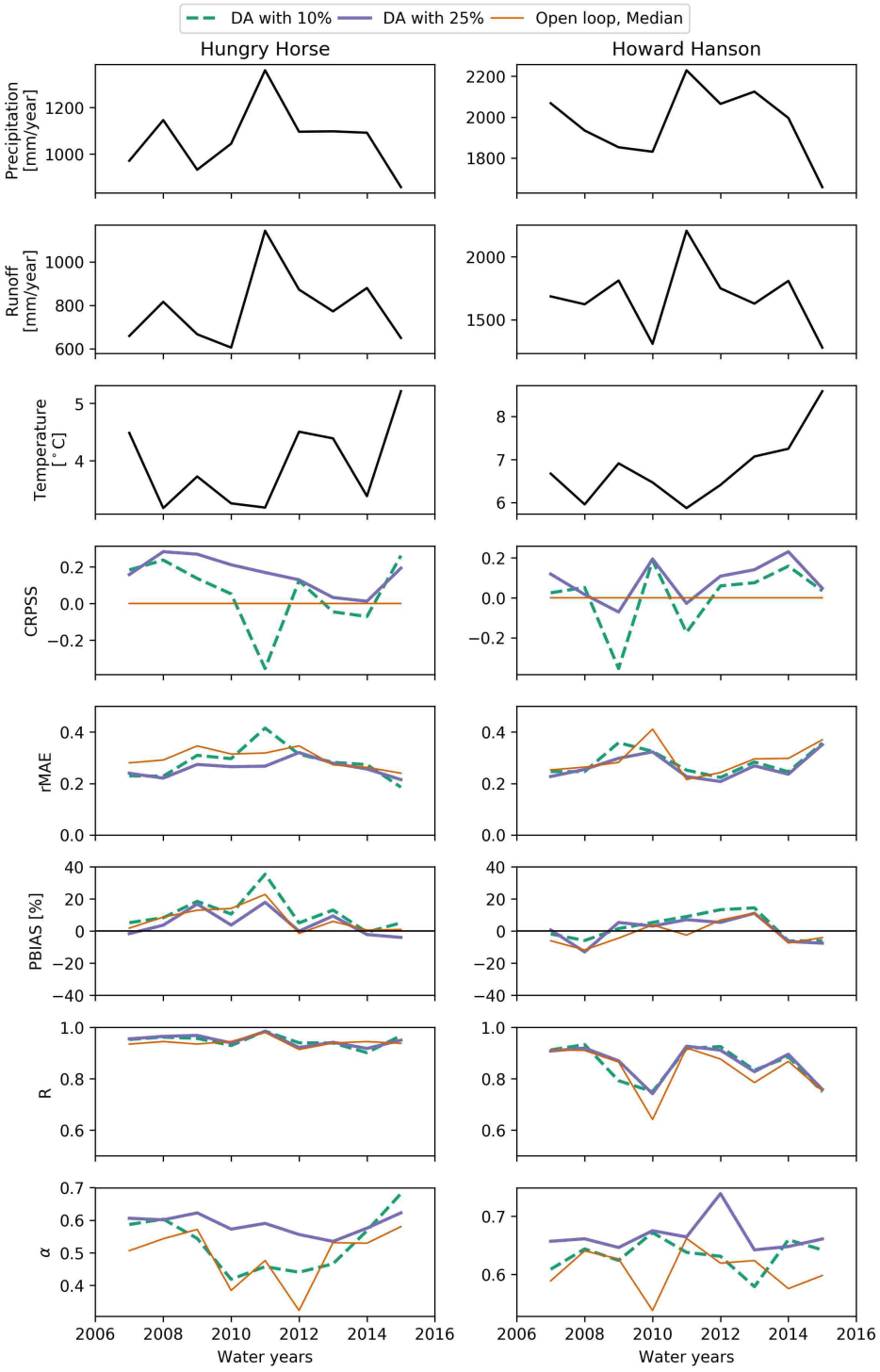

Figure 3.9. Annual time series of total ensemble mean observed precipitation, annual average observed runoff, observed average ensemble mean temperature, CRPSS, Pearson's R, α, PBIAS, and rMAE for Hungry Horse and Howard Hanson for water years 2007-2015. All metrics calculated at 2-day lead time initialized from five IHC ensemble members.

3.4 DISCUSSION

Hindcast experiments allow us to evaluate decisions made in an automated forecast system. First, we evaluate the number of ensemble members needed to initialize streamflow forecasts in a given basin and forecast system (including hydrologic model and source of meteorological forecast). In the case of the NWS models at Hungry Horse and Howard Hanson, we found that regardless of the uncertainty level assumed in the DA, no more than five IHC ensemble members was necessary to initialize streamflow forecasts forced by GEFS data. The resultant 55 streamflow forecast simulations required is likely to be more feasible than performing all 1100 streamflow forecast simulations. Overall, this result was consistent across seasons and for all metrics.

Too few IHC ensemble members underestimate the ensemble spread relative to observed behaviour, even with the uncertainty represented in the GEFS ensemble; therefore, an ensemble with too few members is less likely to capture the observations. Adding more members allows the ensemble mean to better match the observations, which improves PBIAS, MAE, and R. CRPS rewards forecasts that both match the ensemble mean and have a small spread; this means that there is a trade-off between gains in accuracy of the ensemble mean and the number of ensemble members. Because the ensemble mean does not match the observations better with more members (as shown by MAE and PBIAS, which worsened for large numbers of ensemble members), CRPS for the DA cases does not improve with more ensemble members. α continued to improve, but more slowly in the DA cases, for more than five IHC ensemble members; however, getting that initial distribution that is likely to contain the observations (i.e., the closest match to ensemble mean) produced the biggest improvement in reliability as well. In the open loop case, the median ensemble members are used without any information about observed

streamflow. As the number of IHC ensemble members increases, the ensemble becomes more likely to contain the observation, thereby improving its probabilistic metrics. The open loop case outperforms the DA with 10% uncertainty case for high numbers of IHC ensemble members because the DA with 10% uncertainty produces an overconfident forecast, assuming less error than is actually present.

Forecast performance relative to the open loop case depends on the level of uncertainty used in the likelihood function (Eqn. 3.2). DA with 25% uncertainty outperformed DA with 10% uncertainty in all forecast verification metrics in both study basins. This happens because the higher uncertainty allows a larger range of particle values to maintain weights high enough to avoid frequent resampling. The broader spread of the ensemble means that its chances of containing an observation at any given time are higher. If the observation falls inside the ensemble spread, the PF can more easily adjust ensemble weights to track biases like that shown in Fig. 3.3c.

Of the metrics considered here, correlation of the ensemble mean streamflow forecast with observed streamflow is the least sensitive to the initialization method. This suggests that the correlation is strongly dependent on the meteorological forecast, and that much of the signal in the correlation metric is determined by regime or climatological variability, which the models capture fairly well. The other deterministic metrics (PBIAS and MAE) improve with DA, particularly if the uncertainty in the likelihood function is 25%. This means that DA improves the ensemble mean in most cases, particularly at short lead times. The impact of DA on these terms decreases as lead time increases because the forecasts are less sensitive to IHCs at longer lead times. At the shortest lead time, the percent improvement in CRPS is 35% and 19% for

Hungry Horse and Howard Hanson, respectively; and 66% of forecasts were improved for both CRPS and MAE at Hungry Horse, and 64% for Howard Hanson.

At both study sites, reliability, as measured by α, is the metric most improved by DA (and therefore most sensitive to IHCs) at all lead times and in all seasons except for summer. In Chapter 2, we found that DA most improved simulated SWE during the spring (snowmelt) and fall (snow accumulation), and at Hungry Horse. During these dynamic snow seasons, the open loop forecasts are particularly unreliable, suggesting that the model simulations without DA do not adequately represent snow storage dynamics during these dynamic periods. Improved SWE states are most likely responsible for the improvements in all metrics during the spring and the larger improvements in forecasts during spring and summer at Hungry Horse relative to Howard Hanson.

At Howard Hanson, PBIAS is negative at all lead times and in all seasons, and at Hungry Horse, PBIAS is positive in most cases. The time series (not shown) of PBIAS at Howard Hanson shows that most of the negative bias comes from a few isolated events during which the meteorological forecast misses high precipitation. The impact of these missed events is largest at longer lead times (Fig. 3.5) because GEFS forecasts are more accurate closer to the event. DA with 25% uncertainty reduces the magnitude of PBIAS in both basins during the spring, summer and fall, but at Hungry Horse during winter, DA, regardless of uncertainty level, degrades PBIAS. At Hungry Horse, winter streamflow is primarily generated by baseflow processes, while winter floods occur at Howard Hanson. Because the flows are more dynamic at Howard Hanson, the PF is better able to update the IHCs in winter than at Hungry Horse. At Hungry Horse, if the full ensemble has a large enough positive bias at the beginning of winter that the observation is lower than any ensemble members, no amount of resampling can create a

distribution that contains the observation. Because flows are not dynamic during the winter at Hungry Horse, the likelihood of increasing the ensemble spread to encompass the observed flow is low.

Although the results averaged over all nine years for 2-day lead time show an improvement in PBIAS, CRPS, and α for DA with 25% uncertainty and very little difference in these terms relative to the open loop case for DA with 10% uncertainty at both basins (Fig. 3.5), the relative changes due to DA vary from year to year (Figs. 3.7-3.8). If we had only considered water year 2011 for this experiment, we would conclude that the DA with 10% uncertainty performed worse than the open loop case in at least 3 metrics in both basins. 2011 was the wettest water year in both basins, which suggests that the NWS models struggle to estimate high flows, or that the GEFS reforecasts miss the timing or magnitude of large precipitation events. Either way, it is not representative of the other eight years. A longer period of analysis, with greater variability in climate, would be needed to attribute hindcast performance to specific climate patterns.

3.5 CONCLUSIONS

Automated DA methods perform critical adjustments to modeling system errors in real-time. They represent a key pathway toward skillful 'over-the-loop' hydrologic forecasting systems in which human forecasters are not required to make these adjustments. Automated DA also allows for reproducibility and scalability in monitoring and prediction systems. This enables the systematic testing of different implementation choices, as well as the generation of sufficient hindcasts to support robust forecast performance system evaluation. Research is still identifying key strategies for automated DA, which motivates this study assessing a number of choices that are necessary in implementing DA in a forecast system. If the SIR-PF is used, the forecast

system designer must decide how many particles to evolve, which likelihood function to use and how much uncertainty to assign it, how frequently or under which conditions to resample, and how to match IHC ensemble members with meteorological forecast ensemble members. In this study, we examine the uncertainty in the likelihood function in a SIR-PF and how to pair IHC ensemble members with meteorological forecast ensemble members. This study demonstrates the value of hindcasts as a tool to inform these choices before implementation in an operational system. We also examine the implications on 0- to 6-day lead time streamflow forecasts of using SIR-PF to update IHCs in two basins with seasonal snow. The primary findings of our hindcast experiments are as follows:

- Because SIR-PF does not perturb model states, when the full ensemble is completely biased relative to observed streamflow, SIR-PF cannot produce an ensemble that contains the observed streamflow. At best, it can give higher weight to the least biased particle. Future work should consider potential avenues to correct for this bias.
- No more than five IHC ensemble members are required to initialize streamflow forecasts when pairing SIR-PF results with GEFS forecasts in the two basins studied here. Further investigation is necessary to test the robustness of this result to hydrologic model, the number of particles evolved by the SIR-PF, meteorological forecast, and representation of IHC uncertainty (this study only considers the propagation of forcing uncertainties).
- DA with 25% uncertainty in the likelihood function (Eqn. 3.2) outperforms DA with 10% uncertainty in all five forecast verification metrics at both study basins.
- Improvements in forecast accuracy (CRPS, PBIAS, MAE) due to DA decline as lead time increases because the impact of IHCs declines over time; however, improvements in reliability (α) are consistent across lead times.

- SIR-PF of streamflow observations in basins with seasonal snow improves forecast performance in terms of accuracy (CRPS, PBIAS, MAE) during the spring and summer and in terms of reliability (α) during spring and fall. This is consistent with improved SWE estimates during seasons in which snow melt contributes to streamflow (spring and fall).
- Similar to forecast performance, improvements due to DA vary from year-to-year, and the difference between DA with 10% uncertainty and DA with 25% uncertainty also varies from year-to-year. Hindcast studies based on short periods of time could easily over- or under-estimate the improvement in forecast performance due to DA.

Based on these findings, we recommend using hindcasts to test for the appropriate level of uncertainty in a specific hindcast setting. It is likely higher than the true observational uncertainty because of modeling errors. Because of high interannual variability in performance, it is necessary to test these hindcasts over several years. We also recommend using hindcasts to determine the minimum number of IHC ensemble members required in the initialization of streamflow forecasts for a specific system design. Because of the potentially high additional computational and I/O costs of running more ensemble members in the forecast system and the need for extremely fast turn-around from data retrieval to useful real-time streamflow forecast, it is reassuring to find that the number of IHC ensemble members used to initialize forecasts can be smaller than that needed to represent the IHC distribution. For forecast systems in basins with seasonal snow cover, we recommend relying on DA more during spring when snow is actively contributing melt to runoff. This should improve forecasts in both the spring and summer in basins, like Hungry Horse, that have mostly winter precipitation. In basins that receive a mixture

of rain and snow, DA is appropriate during the most dynamic seasons, such as spring and fall at Howard Hanson.

Chapter 4. A CASE STUDY OF ANALOG RESAMPLING FOR PARTICLE FILTER DATA ASSIMILATION IN HYDROLOGIC STATE ESTIMATION

This chapter is in preparation as

Clark, E. A., A. W. Wood, B. Nijssen, and M. P. Clark. A case study of analog resampling for particle filter data assimilation in hydrologic state estimation.

Abstract

Streamflow forecasts depend on the quality of simulated hydrologic state estimates used to initialize a forecasting model, including snow water equivalent and soil moisture. Because of the uncertainties in hydrologic simulations, several studies have investigated the potential for improving estimated states by incorporating observations of hydrologic variables. The Sequential Importance Resampling Particle Filter (SIR-PF) is one approach to assimilate observations. It is attractive to hydrology applications because it works in nonlinear and non-Gaussian systems and also because it does not require the perturbation of states. The primary limitations on the SIR-PF are sample degeneracy and its consequence, sample impoverishment. Sample impoverishment occurs when the ensemble no longer contains enough diversity in states to accurately represent the true distribution of states. Existing approaches to limit sample impoverishment involve state perturbation, such that there is no longer a guarantee that the simulated streamflow, soil moisture and snow water equivalent could co-exist given their shared meteorological history. We present an alternative approach to SIR based on historical analogs to inflate sample spread without perturbing individual states. We demonstrate this approach — termed "Analog Resampling"

(AR) — for particle resampling based on streamflow observations during five hydroclimatically diverse events. Based on the behavior of AR compared to that of SIR-PF, we suggest several lines of future development, including: 1) evaluation of AR-PF methods in a continuous system; 2) evaluation of the impacts of increased ensemble spread due to AR-PF on streamflow forecast reliability; 3) assessment of the impacts on forecast performance of sample correlation; and 4) identification of methods to account for shifting hydroclimate and interannual variability.

4.1 INTRODUCTION

Because measurements of snow water equivalent (SWE) and soil moisture are scarce, hydrologic models are widely used to estimate these conditions for use in initializing streamflow forecasts. Hydrologic models are imperfect [*Clark et al.*, 2008, 2015; *Haughton et al*., 2016], and the meteorological data that drive them also contain errors [*Chappell et al*., 2012]. Data assimilation seeks to minimize the resulting model moisture state errors by incorporating additional observations. The Sequential Importance Resampling (SIR) Particle Filter (PF) [*Kitagawa*, 1996; *Doucet et al.,* 2001], described in Chapters 2 and 3, is one approach to data assimilation that has been tested for hydrologic applications [*Moradkhani et al.*, 2005, 2006; *Weerts and El Serafy*, 2006]. The particle filter is attractive for hydrology applications because it works in non-linear and non-Gaussian systems. It also updates the probability distribution of streamflow and hydrologic states (soil moisture, SWE, etc.) without perturbing individual states [*Salamon and Feyen*, 2010]. This means that the resultant model realizations are internally consistent — each simulation of soil moisture, SWE and streamflow contains a realistic combination of conditions given the history of meteorological forcings.

Unfortunately, the particle filter is also prone to an effect known as sample degeneracy, which can lead to sample impoverishment [*Li et al.*, 2014]. Sample degeneracy occurs when the

majority of the particle weights are given to a few particles, such that most particles no longer contribute meaningful information to the posterior distribution. Several resampling methods have been developed to more evenly distribute the particle weight. In some resampling methods, like the SIR, particles are resampled from the discrete distribution of the current ensemble. This process assigns the same state to multiple particles, while it discards other states. The effect of resampling from a discrete population of particles is sample impoverishment, which occurs when nearly all the particles have the same state [*Li et al*., 2014]. Several methods attempt to counteract this effect and add diversity to particle states, but they also destroy the internal consistency of hydrologic states. For example, the process of roughening [e.g., *Li et al*., 2013] adds independent Gaussian noise to the states of each resampled particle, while regularization [*Musso et al.*, 2001] approximates a continuous probability distribution from the discrete distribution of particle weights and then resamples particle states from that continuous distribution. *Li et al*. [2014] review several additional approaches to counteract sample impoverishment, none of which guarantee that individual particles will have internally consistent model states.

In the case of the hydrologic simulations presented in Chapters 2 and 3, we noted that there were periods when the full ensemble of particles was biased relative to the observation. In this situation, the closest ensemble member is given the highest weight, and the sample degenerates. If the sample is impoverished and states are not perturbed, it can take many days for the ensemble to recover, particularly in hydrologically inactive periods. This is particularly true in our system because model state uncertainty is determined by meteorological forcings alone. In this case, if there is no rainfall in any of the ensemble members, particles that share the same state after resampling will continue to produce the same streamflow, unless differences in

temperature between forcing ensemble members are large enough to create a diversity of snow melt or evaporative conditions.

In this paper, we propose an alternative method for particle resampling that expands the set of candidate particles to include simulated states from prior dates, through inclusion of historical analogs. The motivation for the approach—termed analog resampling (AR)—is two-fold. First, sample degeneracy on the current date may arise from forcing errors (e.g., erroneous lack of rainfall) in the current meteorological analysis, yet watershed states and outputs that better match the current observations may have occurred in the past. Second, the possibility of leveraging an enhanced ensemble of past simulations to maintain an effective real-time filter may reduce the dimensionality and computational requirements for an operational PF implementation. Ideally the AR particle filter (AR-PF) will reduce sample impoverishment by drawing on a much larger set of candidate simulations than would be possible in a single real-time ensemble.

Analog methods have been used in precipitation forecasting [*Lorenz*, 1969; *Hamill and Whitaker*, 2006; *Hamill et al*., 2006] and have also been applied to streamflow forecasting [*Crochet*, 2013; *Koutsoyiannis et al.*, 2008; *Yao and Georgakakos*, 2001; *Wood*, 2012]. The AR-PF differs from the recently proposed Analog Data Assimilation [*Lguenstat et al*., 2017], which replaces the forward model (or hydrology model in our application) with an analog forecast model. Here, we propagate states using a physically based hydrologic model. Analog methods are used to replace particles when sample degeneracy occurs. In that case, particles are replaced with an ensemble of historical simulations sampled from periods in which either the observed or the simulated streamflow is analogous to that of present observed streamflow.

Section 4.2 describes the analog approach for particle filter resampling, illustrated here using two possible strategies for analog identification and selection, both based only on streamflow metrics: one in which analogs are ranked based on matching current observed streamflow to both past observed and simulated streamflow and the other based on matching current observed streamflow to past simulated streamflow only. Exploratory results for one watershed are presented for five events representing a range of hydroclimatic conditions in section 4.3. In section 4.4, we discuss the implications of these results for future applications of the AR-PF, and section 4.5 presents some preliminary conclusions and plans for future work.

4.2 METHODS

Chapters 2 and 3 describe the streamflow hindcast and forecast system used in this research. The portion of the forecast workflow addressed in this chapter (Fig. 4.1) is data assimilation in the creation of weighted ensembles of hydrologic states. By comparing Fig. 4.1 to the SIR-PF workflow (Figs. 2.1 and 3.1), we can see that unlike the SIR-PF, the AR-PF requires an independent set of historical simulations to generate a catalog of potential streamflow analogs. The generation of this catalog is discussed in more detail in section 4.2.2. All other aspects of the workflow—including data sources, ensemble meteorological forcing generation, open loop simulations, and the National Weather Service (NWS) model set up—are as described in Chapters 2 and 3.

4.2.1 *Particle filter*

We use two types of particle filters (PFs) in this study: SIR-PF and AR-PF. Both methods are based on the Sequential Importance Sampling (SIS) PF [*Doucet et al.,* 2001], described in

section 2.2.3.2, in which the weights are calculated as a function of the weight at the previous time step:

$$w_t^i \propto w_{t-1}^i \cdot p\left(z_t \middle| x_t^i\right) \tag{4.1}$$

where w_t^i is the weight of particle i at time t, and $p\left(z_t \middle| x_t^i\right)$ is the likelihood of the observation z_t at time t given state x_t^i of particle i at time t. The likelihood is calculated here as a Gaussian function:

$$p\left(z_t \middle| x_t^i\right) = \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{(z_t - h(x_t^i))^2}{2\sigma_t^2}} \cdot \left(\sum_{m=1}^{N_p} \frac{1}{\sqrt{2\pi\sigma_t^2}} e^{-\frac{(z_t - h(x_t^i))^2}{2\sigma_t^2}} \right)^{-1} \tag{4.2}$$

where N_p is the number of particles, σ_t^2 is the variance of the assumed Gaussian distribution, and $h(x_t^i)$ is the state of particle i at time t transformed to directly simulate the observed quantity. In this case, the unit hydrograph portion of the NWS models, which relates model states to streamflow by routing the model's runoff, is the observation operator h(·). In Chapter 3, it was shown that the PF added more value to short-range streamflow forecasts when the likelihood function had a standard deviation of 25% of current flow instead of 10%, so that value is used to calculate weights in this chapter as well. As in Chapters 2 and 3, we substitute Eqn. 4.2 into Eqn. 4.1 to calculate particle weights, which are then normalized over N_p to sum to one. These weights represent the posterior probability of model states.

In Chapters 2 and 3, when the number of effective particles (N_{eff}) falls below $N_p/5$, we use what *van Leeuwen* [2009] refers to as stochastic universal resampling to resample the particles to reproduce the posterior distribution with a set of N_p equally weighted particles. These equally weighted particles are sampled directly from the same ensemble of particles that existed prior to resampling, with a frequency according to their prior weights. As a result, the states of the particles with the highest original weight replace the states of particles with low weights.

This means that some particles in the resampled ensemble have the same states; however, each resampled particle is propagated forward in time with a unique forcing ensemble member. While the unique forcing ensemble members ideally ensure that the ensemble will inflate in the future, there is an immediate collapse of the ensemble onto a smaller set of states. Because hydrologic systems have memory and because they exhibit varying levels of responsiveness depending on meteorological conditions, the time to recovery of sufficient ensemble spread (to accurately represent the probability distribution of states and to maintain filter performance) can be more than a few days. In addition, if the entire simulated ensemble is biased, resampling from the best current states does not improve the ensemble bias.

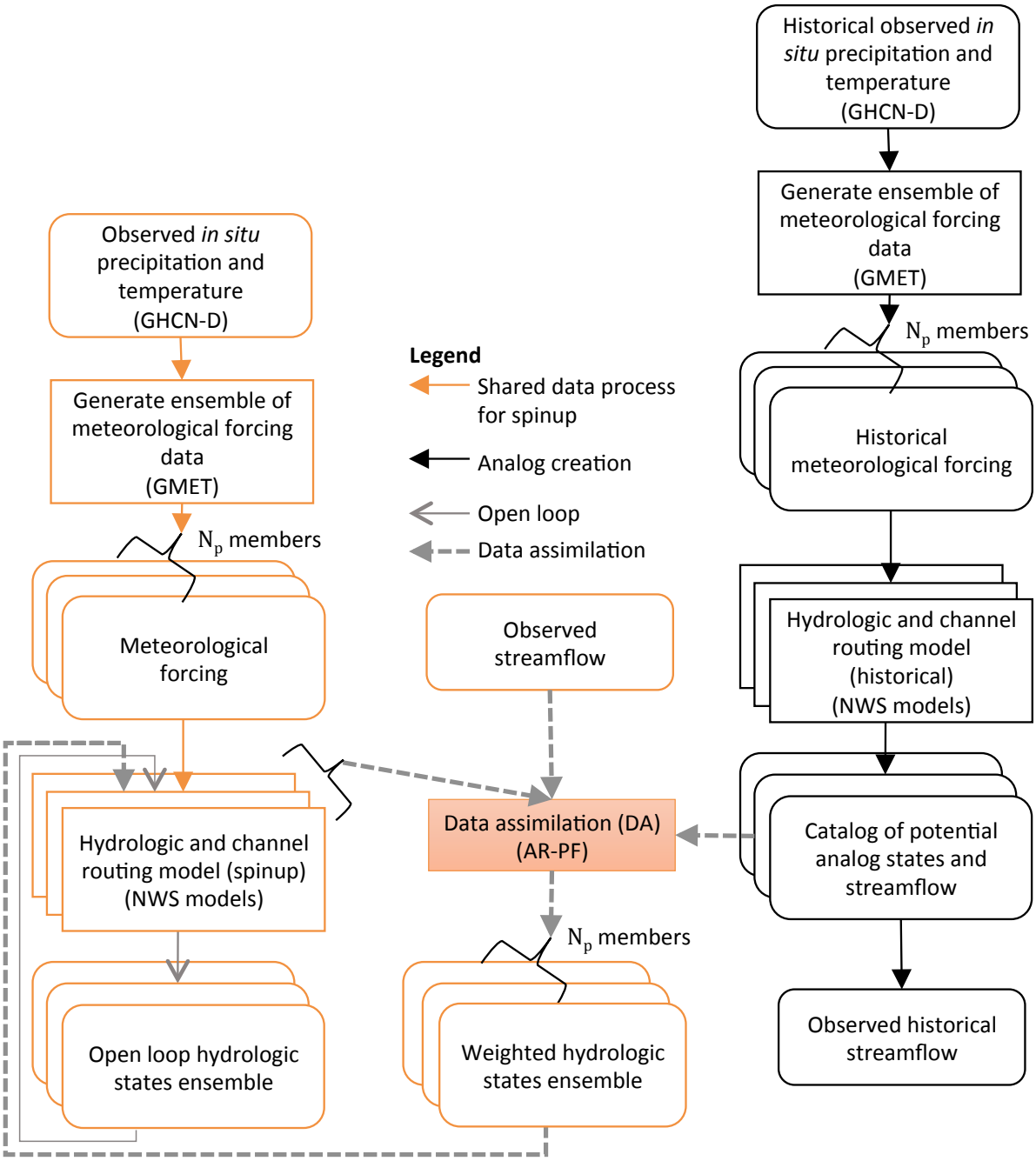

Figure 4.1 Process flow chart for the creation of open loop and weighted hydrologic states ensembles using the Analog Resampling Particle Filter (AR-PF).

4.2.2 *Analog resampling*

The analog approach used in this study replaces NWS model states (soil moisture, SWE, and unit hydrograph history) with NWS model states that have been resampled from candidate simulations taken from an extensive archive (a.k.a. catalog) of previous simulations of hydrologic model states. In a forecasting application, these states can be used as initial conditions for streamflow predictions. In this discussion, the term "analogs" is used to denote the dates and ensemble members in the historical record and the observations and simulations that correspond to those dates and ensemble members. We test the analog approach using two different strategies for identifying and selecting analogs, both of which focus on streamflow rather than additional hydroclimate variables. The "resampling" step of AR is the process of replacing the existing particle states with simulated states from analogous hydrologic conditions in the past.

The first method (termed "obs-sim") compares current streamflow observations (z_t) on day t only to past streamflow simulations ($h(x_{t'}^{p})$) for each day t' (where t' simply denotes a day from the catalog of potential analogs) and potential analog ensemble member p in the catalog. The second selection method (termed "obs-obs") expands the selection criteria to compare present streamflow observations (z_t) both to past streamflow observations ($z_{t'}$) for each day t' in the catalog and to their corresponding streamflow simulations. The former ensures that the analog's simulated streamflow will be a good fit to the current observed streamflow, and the latter ensures that the analog's true hydrologic conditions (as estimated from observed flow) are similar to the current hydrologic conditions. The obs-obs case does not address or target model bias but leaves bias correction for post-processing.

Analog methods are general, but in the temporal application context, they use a look-up procedure and a quantitative metric of similarity (or 'distance') to identify a set of historical

dates on which conditions are analogous to the period of interest, often the present. For both obs-obs and obs-sim selection methods, we construct a catalog of n potential analogs. For the obs-obs method, n is the number of years in the historical data set (26, from 1980 to 2005) times the number of days per year to include in each analysis period (w = window size). For each year from 1980 to 2005, we define the searchable catalog to include the calendar period each year from 15 days before to 15 days after the day of the year of the current analysis date, for a total of w = 31 days. By limiting the analysis period to a window of +/- 15 days, we limit the analog selection to periods in which seasonal hydroclimatic behavior is most likely to be similar to the present. For the obs-sim method, n is the number of ensemble members in the open loop simulation (100) times the number of years in the historical data set (26, from 1980 to 2005) times the number of days per year to include in each analysis (w = window size). For example, if resampling is required on 16 Jan. 2015, the AR-PF will search from 1 Jan. to 31 Jan. for all years in the historical record. Therefore, each time the AR-PF is called, a total of n = 26×31 = 806 and n = 806×100 = 80,600 potential analogs are searched for the obs-obs and obs-sim selection methods, respectively.

To assess each potential analog in the record, we use a similarity metric defined as the Euclidean distance between the time series of analog streamflow and of current observed streamflow over a period of k days up to and including the analog date. The larger the value of k, the more hydrologic memory the system considers in selecting the analogs. This array of potential analogs ($\boldsymbol{z_{t'}} = [z_{t'}, z_{t'-1}, z_{t'-2}, \dots, z_{t'-k-1}]$ for obs-obs or $\mathbf{h}(x_{t'}^{p}) = [h(x_{t'}^{p}), h(x_{t'-1}^{p}), h(x_{t'-2}^{p}), \dots, h(x_{t'-k-1}^{p})]$ for obs-sim) is then compared to the k-dimensional vector of observed streamflow from the past k days ($\boldsymbol{z_t} = [z_t, z_{t-1}, z_{t-2}, \dots, z_{t-k-1}]$). To achieve a computationally efficient ranking of analogs, we use a k-d tree [*Maneewongvatana and Mount*,

1999], as implemented in the scipy.spatial.KDTree python class, to quickly locate the approximate "nearest neighbors" (best matching analogs) in a k-dimensional space. Preliminary tests showed that the resultant analog ensemble was similar for values of k between 7 and 10 days (not shown). The results shown in this chapter use k = 7 days.

The obs-sim analog selection directly returns 100 nearest neighbors (analogs) that replace the current 100 particles at time t. Because the catalog in the obs-sim case contains 100 particles for each day t', the number of unique dates from which the analogs were selected is usually under 100. For example, in the five cases described in section 4.2.3, the largest number of unique dates returned by the obs-sim look-up was 55.

The obs-obs analog selection method works in two steps. Each day t' has an ensemble of 100 simulated flows from the archive that correspond to only one observed flow. This means that if the analog search returns 100 nearest neighbors for the obs-obs method, there would be 100 distinct dates, with a corresponding total of 10,000 potential analogs (each analog corresponding to a date and an ensemble member). Initial investigations (not shown) suggest that subsetting the 10,000 analogs to only include the 100 realizations for which $h(x_{t'}^{p})$ best matches $z_{t'}$, we can end up with many analogs for which simulated streamflow $h(x_{t'}^{p})$ is a poor match to the current observation z_t. To reduce the search range and still satisfy the goal of the obs-obs selection method — identifying analogs for which the true hydroclimatic conditions are similar at those of the date of analysis – we instead select only the 10 nearest neighbors in the obs-obs look-up. These 10 dates provide a total of 1000 potential analogs. The final 100 analogs are selected from this set as those that minimize the mean absolute error between observed analog and simulated analog streamflow for the k (7) days leading up to and including the analog date.

The catalog is constructed following a similar process as that used in generating the open loop simulations, described in sections 2.2.3.1 and 3.2.3.1. In situ observations of precipitation and minimum and maximum temperature are obtained from the Global Historical Climatology Network-Daily dataset [*Menne et al.*, 2012a, 2012b] for the period 1 Jan. 1970 to 31 Dec. 2005. After quality control and gap filling, the observations are probabilistically interpolated to an ensemble of 1/16° gridded meteorological time series using the Gridded Meteorological Analysis Tool (GMET) [*Newman et al.,* 2015]. The time series are then conservatively remapped to the hydrology model's hydrologic response units. These are used to force the same hydrologic model configuration used in simulating initial hydrologic conditions and performing forecast simulations (Fig. 4.1, sections 2.3.2 and 3.2.3). We use the period from 1 Jan. 1970 to 31 Dec. 1980 to "spin-up" hydrologic states. Each of the 100 ensemble members and date from 1 Jan. 1980 to 31 Dec. 2005 is a unique potential analog.

4.2.3 *Case study*

As shown in Chapter 3, compared to the open loop case, when analyzed over nine water years, the SIR-PF contributed limited improvement to medium-range streamflow forecasts at Howard Hanson dam on the Green River at lead times of two or more days. One of the reasons for this lack of improvement is that sample impoverishment is responsible for overconfident forecasts in this basin. Therefore, we selected this site to explore the potential for improvement in initial conditions with a PF using analog sampling.

As an initial exploration of the feasibility of the AR-PF as a strategy for improving the resampling of particles in the context of hydrologic state estimation, we present five examples of historical streamflow events for which the SIR-PF implemented in Chapter 3 required resampling. The event dates were selected to show a range of hydrologic conditions, including:

1) mid-winter freeze (24 Jan. 2009), 2) the rising limb of a spring freshet (14 May 2008), 3) the falling limb of a spring freshet (20 May 2008), 4) the peak of a fall rain-induced event (12 Nov. 2008), and 5) summer-time baseflow (30 Aug. 2015). The purpose of these examples is not to comprehensively evaluate the performance of the AR-PF methods, but rather to serve as an initial assessment of the sensitivity of the approach's performance to hydrologic conditions and analog search and selection method. On each date, we compare the resampled particles produced by the stochastic universal resampling of the SIR-PF with those produced by each method of analog resampling (AR-PF).

4.3 RESULTS

Several features of the results are consistent across the ensembles on all five dates of analysis, regardless of hydrologic conditions (Figs. 4.2-4.6). First, the simulated streamflow from the AR-PF based on an obs-sim lookup best matches the current streamflow observations over the most recent seven days, except in the summer baseflow case for which SIR-PF minimizes the ensemble mean bias over that period (Fig. 4.6). The summer baseflow case is from 2015, which was among the driest years in the simulated history. Second, while the SWE and soil moisture states in the SIR-PF have very little spread after resampling, all AR-PF methods produce larger spread in SWE and soil moisture states than the SIR-PF, and in some cases, the spread in states produced by the AR-PF is greater than that of the open loop. The spread in SWE and soil moisture is increased by the AR-PF relative to the SIR-PF even in cases when the spread in streamflow is reduced by as much as the SIR-PF, relative to the open loop ensemble (e.g., Fig. 4.2).

Fig. 4.2 shows an example of resampled particles during a period of mid-winter freeze and coincident baseflow recession. The open loop model captures the observed flow relatively

well in this case, but the SIR-PF ensemble has a small low bias for several days prior to 24 Jan. 2009 that it was not able to correct with resampling. Both look-up schemes for the AR-PF were able to closely match the observed streamflow. The obs-obs look-up retains slightly more spread in simulated streamflow on the date of analysis; however, it has slightly less spread in soil moisture than the obs-sim look-up. In both AR-PF cases, the low bias in streamflow was corrected, on average, by selecting analogs with lower SWE and soil moisture.

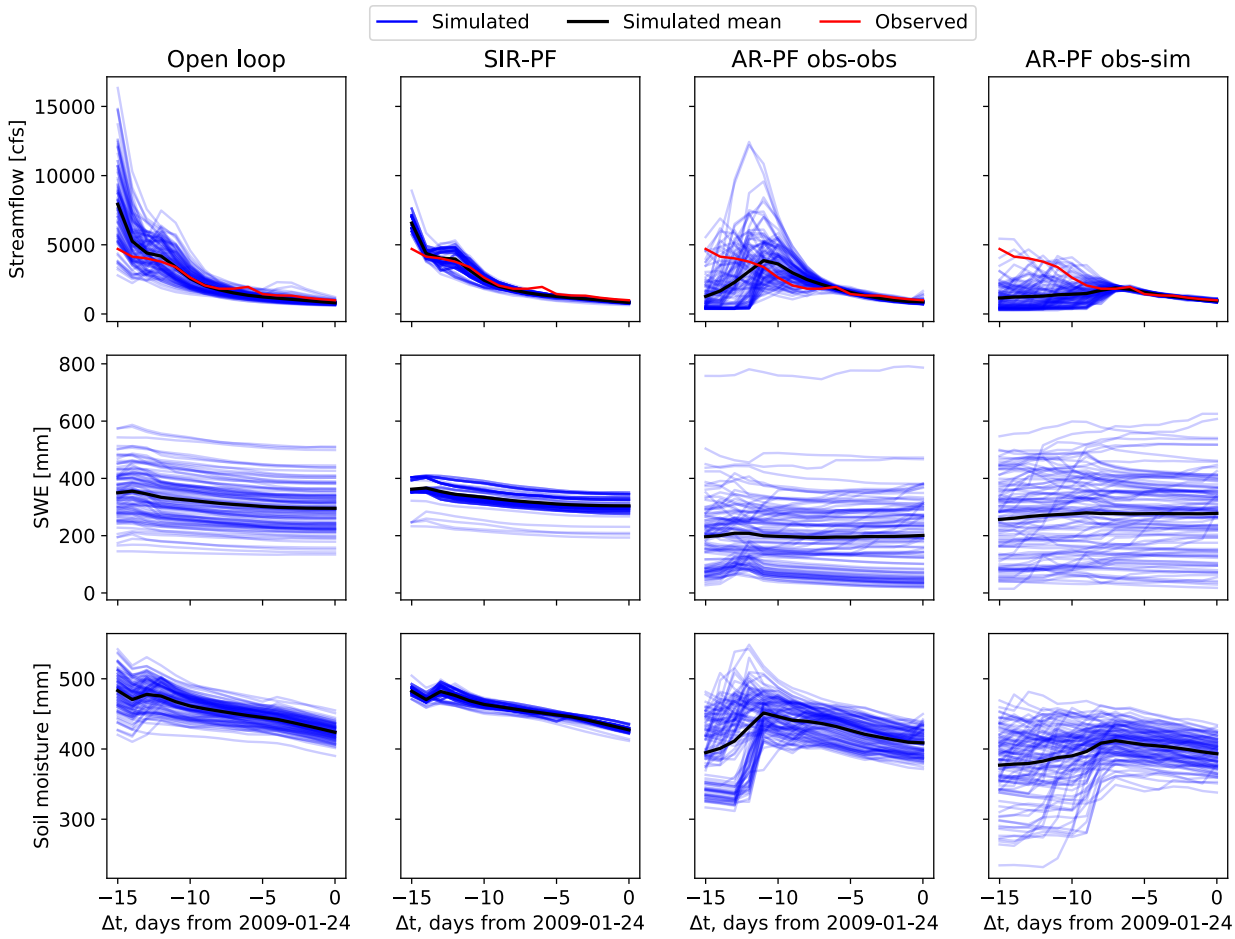

Figure 4.2. Time series of streamflow, soil moisture and SWE leading up to 24 Jan. 2009 for open loop, SIR-PF and AR-PF (both obs-obs and obs-sim look-up methods). For PFs, only the particles that remain after resampling are plotted. The color saturation of the blue lines reflects the density of particles.

Figs. 4.3 and 4.4 show the effects of resampling on the rising (14 May 2008) and falling (20 May 2008) limbs, respectively, of a spring freshet in mid-May 2008. In this case, the only method that captures the observed streamflow in the rising limb is the AR-PF with an obs-sim look-up (Fig. 4.3). It also produces a particle distribution on 20 May 2008 that most closely matched the peak streamflow from three days prior, which could correspond to more accurate state evolution (Fig. 4.4). In this case, the obs-sim AR-PF contains ensemble members whose simulated flow is higher than observed in the analog record, which is why the obs-obs AR-PF underpredicts flows. On the rising limb, all cases result in ensembles with similar rates of snowmelt (the slope of the SWE time series) leading up to the event (Fig. 4.4), suggesting that the best-matching analogs in the obs-sim AR-PF case were taken from analogs that experienced more precipitation than the forcings ensemble on 14 May 2008 (as shown in the difference between open loop flows and AR-PF obs-sim flows). This is an example of the desired AR-PF behavior.

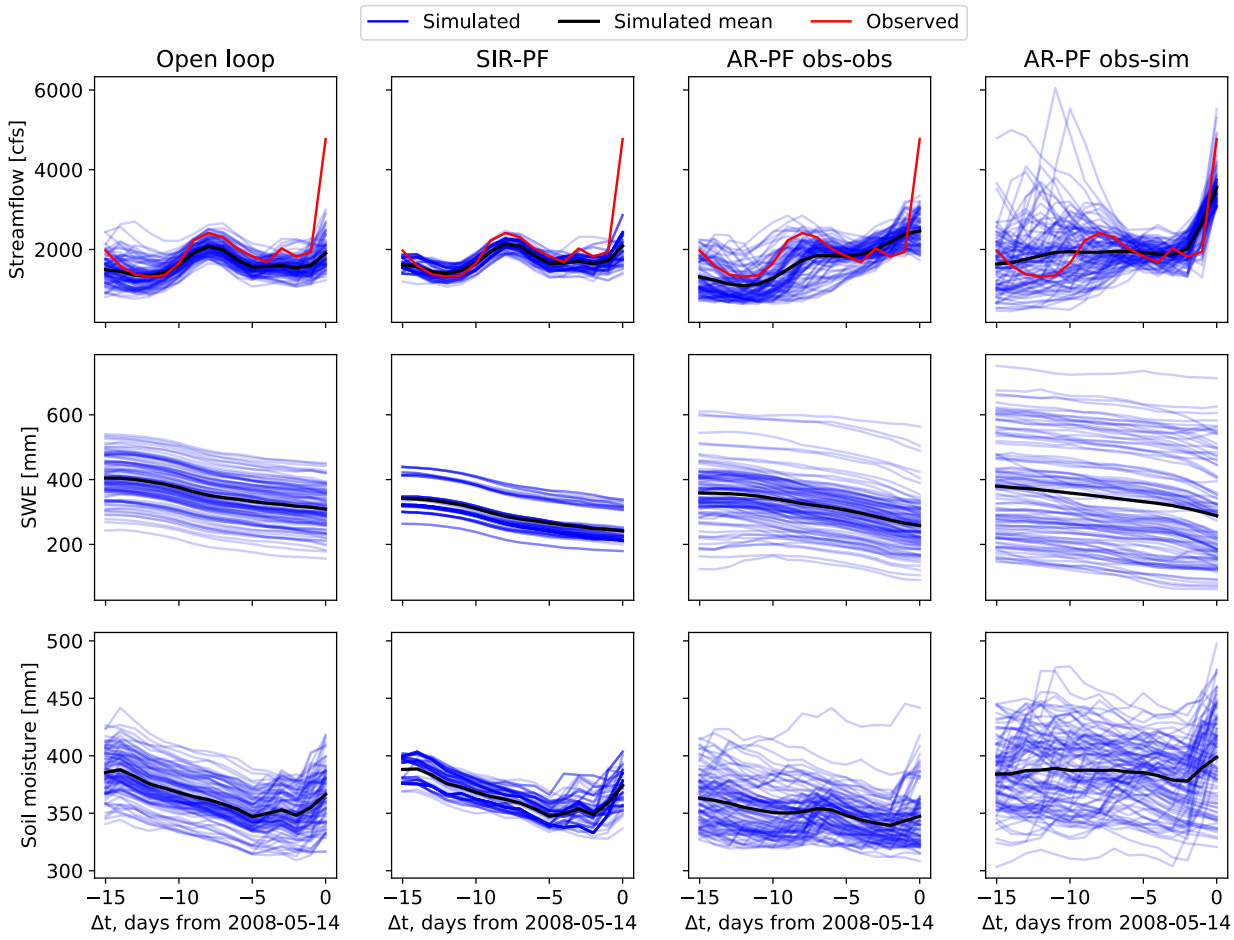

Figure 4.3. Same as Figure 4.2 but for variables leading up to 14 May 2008.

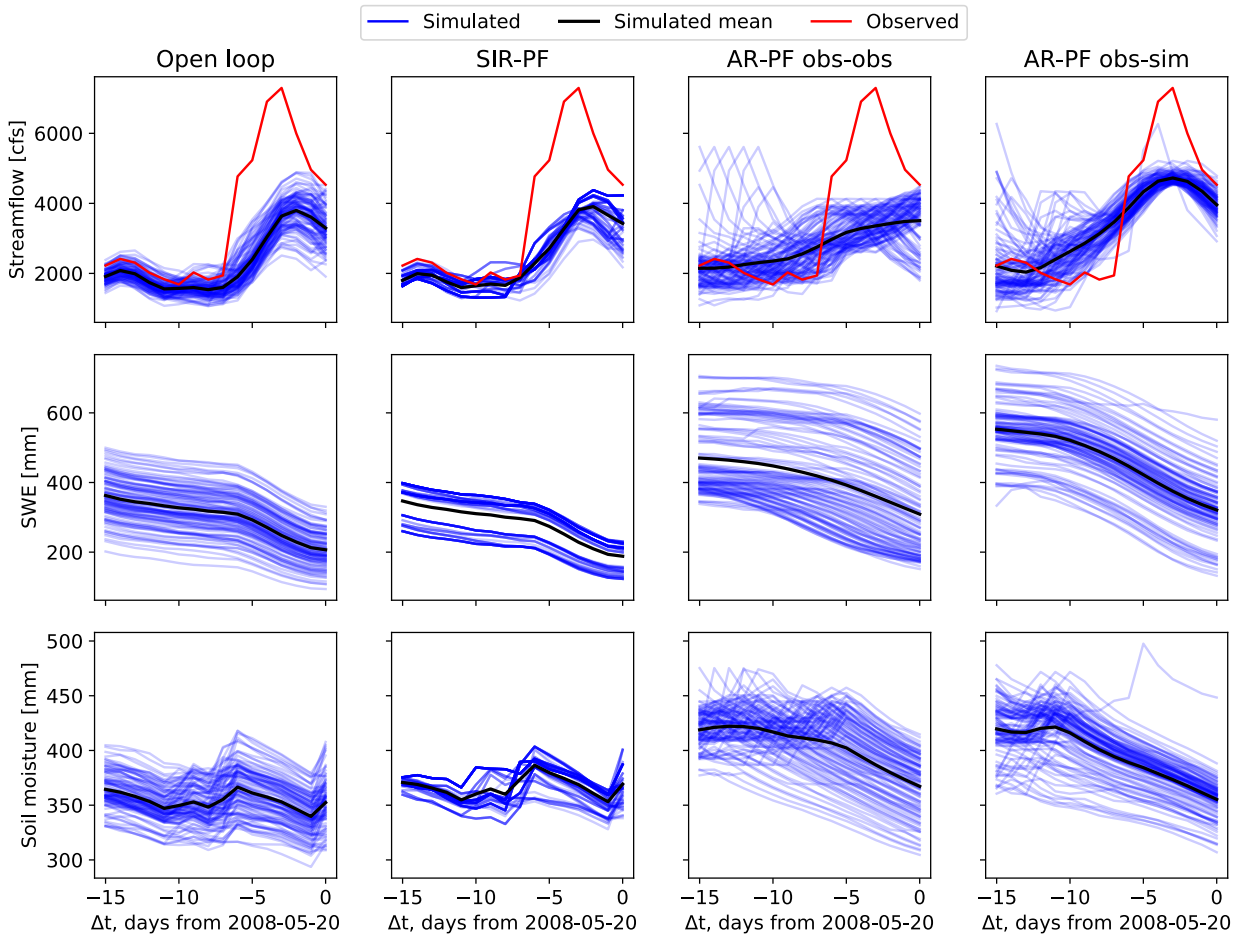

Figure 4.4. Same as Figure 4.2 but for variables leading up to 20 May 2008.

The fourth example shows resampling applied at the peak of a fall rain event on 12 Nov. 2008 (Fig. 4.5). The open loop simulation begins to predict some snow accumulation at the beginning of the seven-day nearest-neighbor analysis window. In this case, analogs were selected from a subset of the historical archive containing 28 Oct. to 27 Nov. of all archived years. The Green River basin frequently has some snow cover at higher elevations during this period. As a result, many of the selected analogs have much more snow than the open loop and SIR-PF predict. 2015 was a particularly hot and dry year, and there would have been very little snow in the basin, if any, on 12 Nov. 2015. Additional simulations are needed to investigate the implications of this erroneous SWE as the ensemble evolves. This suggests that it may be

necessary to further condition the analog selection on factors such as temperature, in addition to time of year.

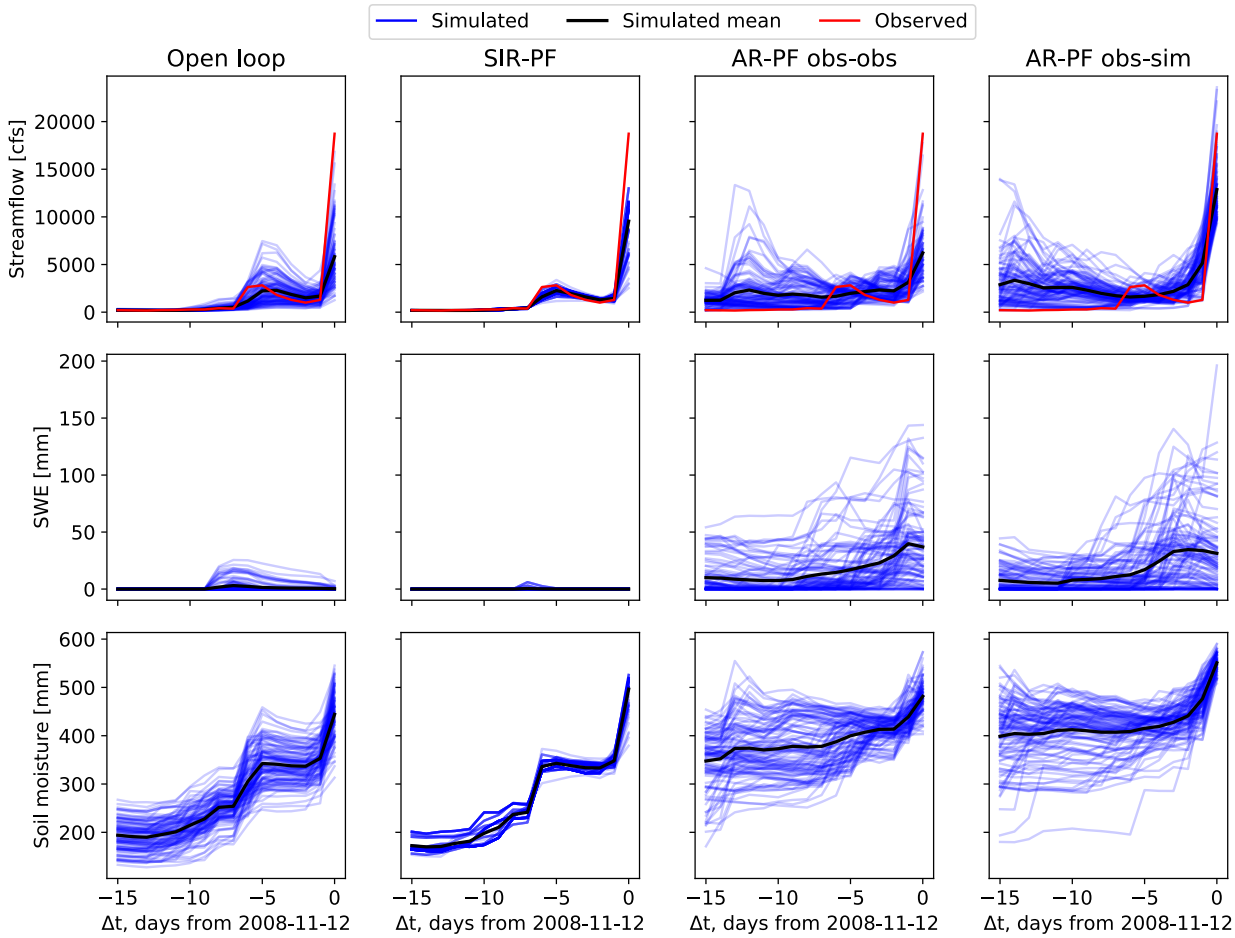

Figure 4.5. Same as Figure 4.2 but for variables leading up to 12 Nov. 2008.

Finally, Fig. 4.6 shows an example of the AR-PF applied to a period of summer baseflow. The date shown, 30 Aug. 2015, is months into a long dry season and represents baseflow only. The open loop simulations, in this case, simulated up to three times as much flow as observed. SIR-PF is effective in reducing the mean streamflow bias; however, the resultant ensemble of streamflow and states is overconfident. Both AR-PF methods reduce the mean streamflow bias relative to the open loop to a lesser extent than the SIR-PF; however, both also retain more spread in both streamflow and soil moisture ensembles, thereby retaining a more

realistic representation of the ensemble uncertainty than the SIR-PF. None of the years from 1980-2005 had flow as low as the observed leading up to 30 Aug. 2015, so even in cases where simulated analog flow matches the observed analog flow well, both simulated and observed analog flows are too high to represent the 30 Aug. 2015 streamflow. This highlights a weakness of this and all analog approaches, which is that the quality of the analysis depends on the archive containing hydroclimatic extremes.

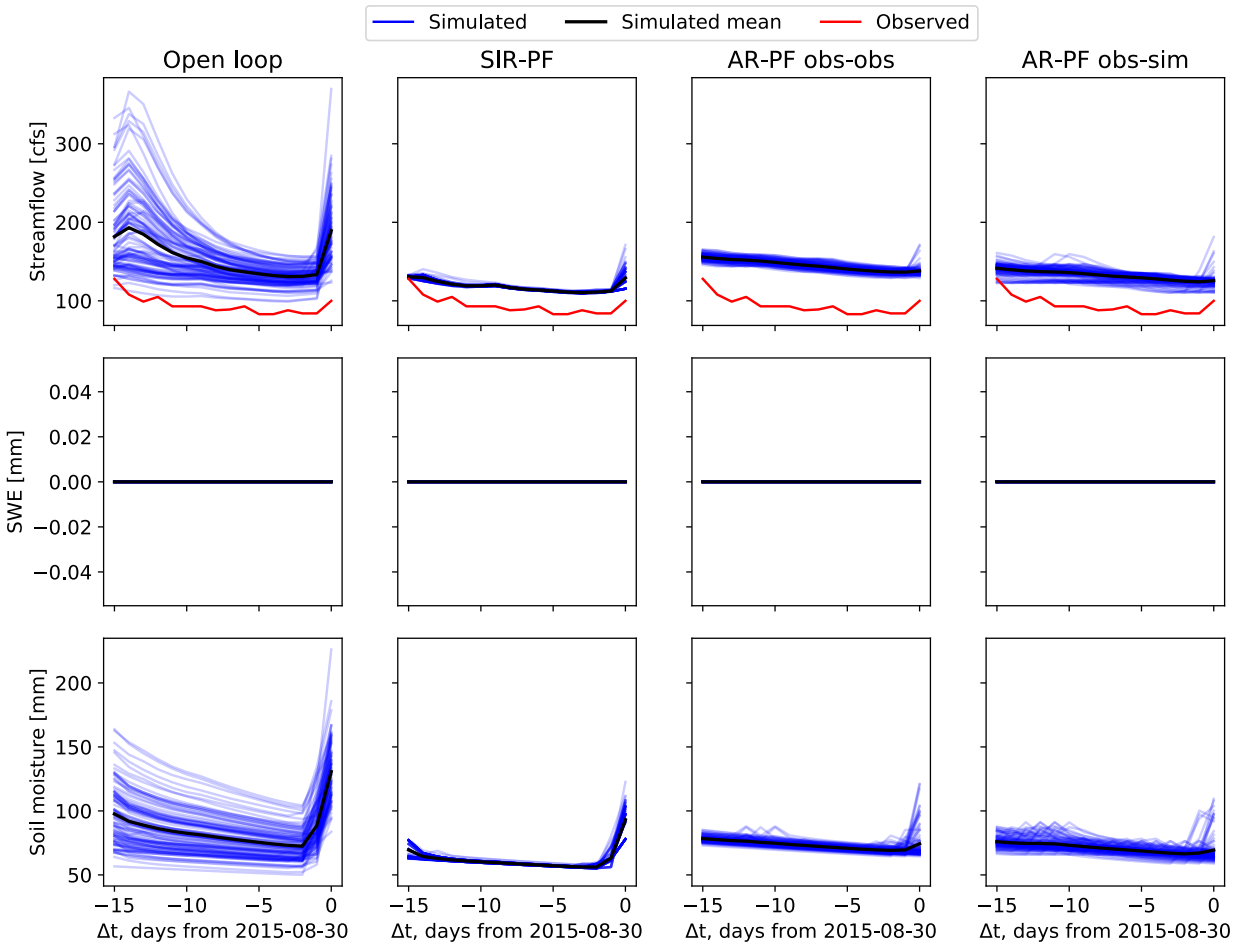

Figure 4.6. Same as Figure 4.2 but for variables leading up to 30 Aug. 2015.

To better understand the source of differences in ensembles generated by the obs-obs and obs-sim analog look-up methods, we examine the observations corresponding to the selected analogs for each analysis date (Fig. 4.7). This examination illuminates two idiosyncrasies that may arise from the current methodology: the selected analog set can contain 1) several sequential days from the historical record, appearing as a phase shift, or 2) several ensemble members from the same day in the historical record, appearing as a thicker, less transparent blue line in Fig. 4.7. Both effects occur for both methods. The phase shift can also be seen in the simulated flow ensembles, as for AR-PF obs-obs in Fig. 4.4. Note that the existence of analog ensemble members differing only in phase may be seen as undesirable mainly because they contain a structure that does not arise from the uncertainty generation methods and assumptions used to create the open loop states, and also because they introduce a correlation between ensemble members. It is not clear whether this structure would negatively impact the use of the states to initialize forecasts.

The other notable feature of Fig. 4.7 is that the observations corresponding to the selected analogs for AR-PF obs-sim can be quite different from the observations leading up to the day of analysis. This is particularly true for analogs sampled on 30 Aug. 2015. The historical ensemble of gridded meteorology missed a rain event on 29-31 Aug. 2005, resulting in a simulation of continuous low flow. This simulation was sampled 56 times in the obs-sim case. In this case, the analog simulation conditions are still a reasonable approximation to the states corresponding to current observations because in both cases, there is no additional rainfall. It is possible that the 2005 event would have been selected in the obs-obs case if more than 10 analog dates were sampled from the catalog. For the case of 14 May 2008, on the other hand, the wide range of analog observed streamflow for AR-PF obs-sim is indicative of less hydroclimatic similarity

between analogs and the present. This is reflected in the wider range of analog simulated streamflow leading up to the seven-day matching window and the wide range of analog simulated soil moisture and SWE states (Fig. 4.3). These results are consistent with the design of the two analog search approaches, in that the obs-sim analog search approach was not constrained to achieve similarity with past observations, in contrast to the obs-obs approach.

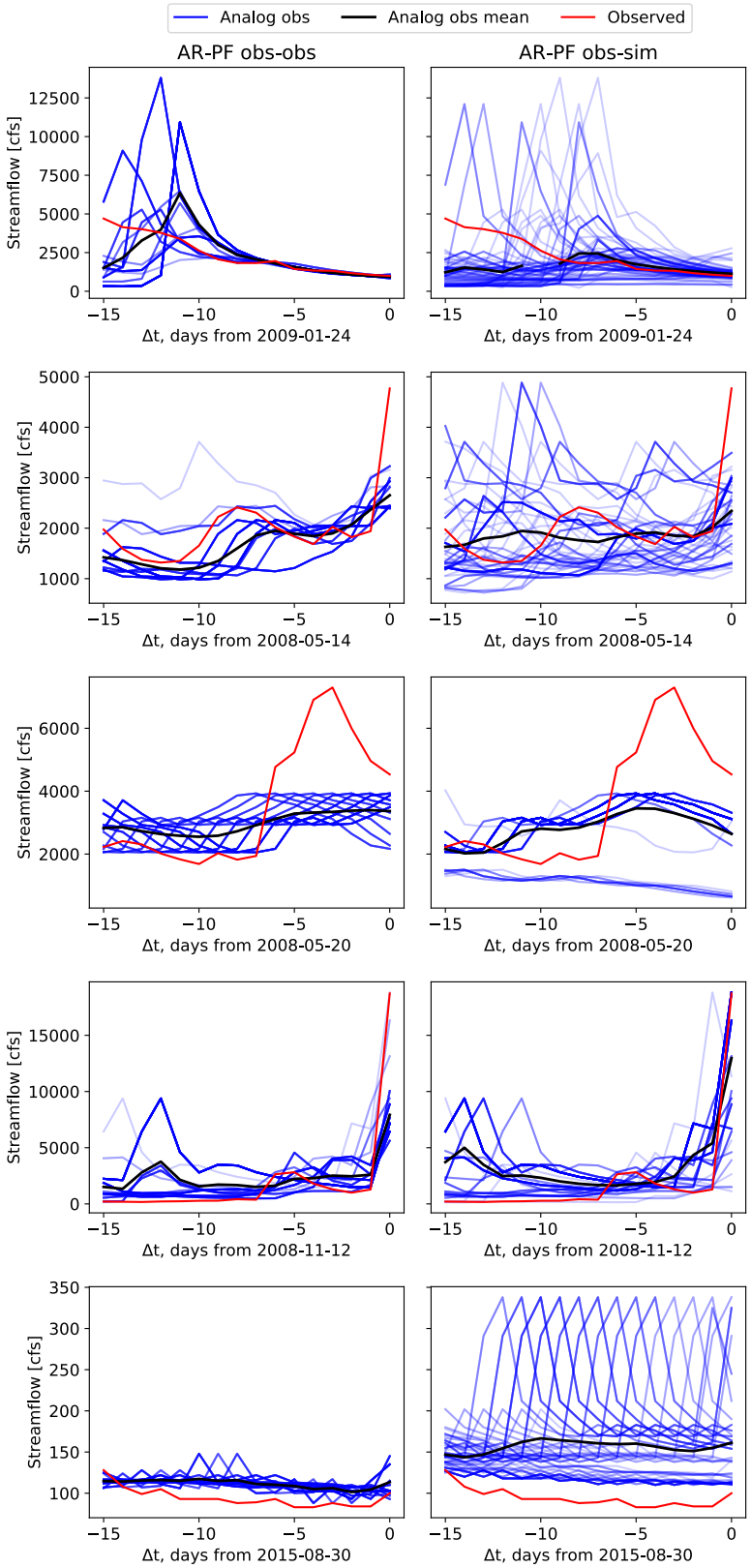

Figure 4.7. Observed streamflow corresponding to the analogs whose simulated flows are shown in Figs. 4.2-4.6. The color saturation of the blue lines reflects the density of particles.

4.4 DISCUSSION

The preliminary analysis presented in this chapter shows that AR-PF has the potential to expand the spread of ensemble states in a collapsed particle filter while maintaining model simulation traces that are internally consistent (by not perturbing individual states). While these individual examples cannot be generalized to assess overall filter performance, they do show promise in addressing a prominent PF challenge, particularly in the case of AR-PF obs-sim. Still, several aspects of the AR-PF strategy and possible approaches to account for them are worth noting in more detail. These include: 1) sensitivity to hydroclimate, 2) system memory, 3) accuracy and reliability trade-offs, and 4) sample independence.

First, because hydroclimatic processes are often tied closely to the time of year in which they occur, the AR-PF is able to efficiently search comparable past seasons to find analogs that reproduce current observed streamflow as well as, and often better than, the SIR-PF. In a changing climate, the window for analog identification might need to shift or expand. For example, in basins with seasonal snow in Washington State, a shift from snow to rain as the primary form of precipitation is projected in the coming decades [*Elsner et al*., 2010]. The example for 12 Nov. 2008 (Fig. 4.5) shows the sort of mismatch in conditions that we might expect under a changing climate, where a flood event driven by heavy rains in the absence of snow is replaced by states associated with rain-on-snow flooding. If a time of year during which there had historically been snow on the ground—like late October—ceases to have snow on the ground, more appropriate analogs may be found at other times of year, for example by looking for analogs to October observations in September or by defining the search window based on process rather than time. Under a stationary climate assumption, the use of a search window

indexed to the day of the year or season is essentially a proxy for an explicit process-based analog search.

Second, it is possible to improve the simulation of streamflow over a period of 7 days without improving the simulated states. The inherent equifinality of hydrologic models, which implies that similar runoff can arise from different model state and parameter configurations, cannot be easily overcome [*Beven*, 1993], but the degree to which poor analog states are selected can be minimized. The effect of improving streamflow while degrading hydrologic states is apparent in Fig. 4.5. The best analogs in terms of simulated streamflow for this event also have significant amounts of snow at a time when no snow was present. Interannual variability in hydroclimate can lead to streamflow analogs with significantly different states, some of which we know to be wrong. Some possible approaches to limit these differences include extending the memory (k), or basing the catalog on process rather than time, to the extent that watershed processes can be readily recognized. An obvious path toward achieving greater fidelity in analog selection is to expand the hydroclimate feature vector used to calculate analog similarity to current conditions. Here we present just two possible analog similarity calculations, both of which considered only streamflow (e.g., observed and simulated), but observations of SWE and soil moisture, as well as basin meteorological data, where available, would likely be very effective in further constraining analog selection. Nonetheless, one benefit of the AR-PF is that it increases the spread of possible particles at the time of resampling such that the filter does not become fixed on an erroneous path, tracking the evolution of only a few (potentially wrong) states. The increased spread increases the chances of the distribution containing reasonable state estimates, and the most reasonable states (particles) may be preferentially selected when their

linkage to streamflow emerges (e.g., unreasonable SWE states may be de-ranked when their melt appears as inconsistent to flow).

Third, in ensemble forecasting, there is often a trade-off between mean error of the ensemble members and reliability. After the SIR-PF resamples the particles, the mean error is often decreased because the ensemble only contains particles closest to the observations. However, in dropping the poor-performing states, the ensemble becomes overconfident (the spread is too small). This trade-off is apparent in Fig. 4.6, for which the ensemble mean of the SIR-PF is closest to the observations. While the AR-PF often reduces model bias, particularly for the obs-sim AR-PF, the primary benefit of the AR-PF is that it re-inflates the ensemble spread, resulting in higher reliability. This happens even in cases where the ensemble bias is higher for AR-PF than SIR-PF (e.g., Fig. 4.6).

Fourth, as seen in Fig. 4.7, the AR-PF sometimes selects several ensemble members from the same day or several sequential days from the historic record as analogs. This raises the concern of sample independence. Each member of the gridded meteorological forcing ensemble is an independent sample of the meteorological forcing distribution [*Clark and Slater*, 2006; *Newman et al*., 2015], which means that the hydrologic ensemble members in the open loop case are also independent. Therefore, the selection of several ensemble members from the same day should not impact the statistical independence of the resampled ensemble. Sequential days from the same ensemble member, however, will have correlated errors. This effect cannot be entirely removed without limiting each historical ensemble member to contribute no more than one analog date to the resampled ensemble. Because hydrologic states tend to reset in October in this basin, future applications could apply this limit to allow no more than one analog date per water year per historical ensemble member. Because the uncorrelated forcings applied to each state

cause the members to diverge over time, this correlation may prove to be unimportant in practical applications.

4.5 CONCLUSIONS AND FUTURE WORK

An exploratory analysis of the use of an AR-PF approach to replace particles when the particle filter collapses (number of effective particles falls below a minimum) suggests that AR-PF has the potential to increase the ensemble spread following resampling while also improving ensemble mean streamflow estimates. Both of the analog selection strategies demonstrated here increase the ensemble spread of resampled hydrologic states relative to the SIR-PF. AR-PF also has the potential to reduce the computational cost of PFs in operational systems by replacing the simulation of ensemble members with the selection of analogs from past simulations. The AR-PF obs-obs expands the analog search to consider both simulated and observed streamflow and illustrates that constraining the analogs to more closely match current hydroclimate (as indexed by streamflow only) may produce a better ensemble of multiple hydrologic states, which could ultimately lead to better forecasts. The relatively better performance of the AR-PF obs-sim analog search in matching observed streamflow, on the other hand, results in less biased ensemble of streamflow during the spin-up period, which could lead to reduced streamflow forecast errors at initial lead times. To more effectively assess the utility of each of these methods, future work will include a comparison of streamflow forecasts initialized with analog states from both search strategies used here, and possibly use additional variables in the analog search and selection.

The case study results presented here suggest four areas for further development and investigation. These include:

1) Evaluation of AR-PF methods in a continuous system (rather than for specific events).
2) Evaluation of the impacts of increased ensemble spread due to AR-PF on streamflow forecast reliability.
3) Assessment of the impacts on forecast performance of sample correlation from the selection of analogs from the same historic ensemble member over multiple sequential days.
4) Identification of methods to account for shifting hydroclimate and interannual variability.

Several examples are not sufficient to demonstrate improvements or to calculate meaningful verification statistics to evaluate the ensemble of hydrologic states. On-going work will extend the AR-PF methods to simulate initial hydrologic conditions and hindcasts for more water years. There are several degrees of freedom in this analysis including the catalog window size (w), the number of days in the nearest neighbor search (k), and the number of analog days to search for in the k-d tree when performing an obs-obs look-up. The appropriate values for these quantities is likely to vary between basins, and a full examination of the sensitivity of results to each is also needed.

BIBLIOGRAPHY

Abaza, M., F. Anctil, V. Fortin, and R. Turcotte (2015), Exploration of sequential streamflow assimilation in snow dominated watersheds, Adv. Water Resour., 86, 414-424, doi: 10.1016/j.advwatres.2015.10.008.

Anderson, E. A. (1973), National Weather Service River Forecast System--Snow accumulation and ablation model, NOAA Tech. Memo, NWS HYDRO-17, 217 pp.

Andrieu, C., A. Doucet, and R. Holenstein (2010), Particle Markov chain Monte Carlo methods, J. Royal Stats. Soc.: B, 72, 269-342, doi:10.1111/j.1467-9868.2009.00736.x.

Arulampalam, S., S. Maskell, N. Gordon, and T. Clapp (2002), A tutorial on particle filters for online non-linear/non-Gaussian Bayesian tracking, IEEE Trans. Signal Process., 50(2), 174-188.

Beven, K. (1993), Prophecy, reality and uncertainty in distributed hydrological modeling, Adv. Water Resour., 16, 41-51, doi:10.1016/0309-1708(93)90028-E.

Buizza, R., P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei (2005), A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems, Mon. Wea. Review, 133, 1076-1097, doi:10.1175/MWR2905.1.

Burnash, R. J. C., R. L. Ferral, and R. A. McGuire (1973), A generalized streamflow simulation system conceptual model for digital computers, U.S. Department of Commerce National Weather Service and State of California Department of Water.

Chappell, A., L. Renzullo, and M. Haylock (2012), Spatial uncertainty to determine reliable daily precipitation maps, J. Geophys. Res., 117, D17115, doi:10.1029/2012JD017718.

Clark, M. P., and A. G. Slater (2006), Probabilistic quantitative precipitation estimation in complex terrain, J. Hydrometeor., 7, 3–22, doi:10.1175/JHM474.1.

Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008), Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44, W00B02, doi:10.1029/2007WR006735.

Clark, M. P., et al. (2015), A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resour. Res., 51, 2498–2514, doi:10.1002/2015WR017198.

The Climate Corporation (2015), properscoring, version 0.1, https://github.com/TheClimateCorporation/properscoring.

Cloke, H. L., and F. Pappenberger (2009), Ensemble flood forecasting: A review, J. Hydrol., 375, 613-626, doi:10.1016/j.jhydrol.2009.06.005.

Columbia River Management Joint Operating Committee (2015), No Regulation, No Irrigation Flows 1928-2008 Corrected 03-2015. Bonneville Power Administration. http://www.bpa.gov/power/streamflow/default.aspx.

Computational and Information Systems Laboratory (2012), Yellowstone: IBM iDataPlex System (NCAR Community Computing). Boulder, CO: National Center for Atmospheric Research. http://n2t.net/ark:/85065/d7wd3xhc.

Crochet, P. (2013), Probabilistic daily streamflow forecast using an analogue sorting method, Icelandic Meteorological Office, Report No. VI 2013-006.

Daley, R. (1991), Atmospheric Data Analysis, Cambridge University Press, New York.

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

DeChant, C. and H. Moradkhani (2011b), Radiance data assimilation for operational snow and streamflow forecasting, Adv. Water Resour., 34, 351–364, doi:10.1016/j.advwatres.2010.12.009.

DeChant, C. M., and H. Moradkhani (2014), Toward a reliable prediction of seasonal forecast uncertainty: Addressing model and initial condition uncertainty with ensemble data assimilation and Sequential Bayesian Combination, J. Hydrol., 519, 2967-2977, doi:10.1016/j.jhydrol.2014.05.045.

Doucet, A., S. Godsill, and C. Andrieu (1999), On sequential Monte Carlo sampling methods for Bayesian filtering, Stat. Comput., 10, 197-208.

Doucet, A., N. de Freitas, and N. Gordon, eds. (2001), Sequential Monte Carlo Methods in Practice, Springer, 581 pp.

Dumedah, G. and P. Coulibaly (2013), Evaluating forecasting performance for data assimilation methods: The ensemble Kalman filter, the particle filter, and the evolutionary-based assimilation, Adv. Water Resour., 60, 47-63, doi:10.1016/j.advwatres.2013.07.007.

Elsner, M.M., L. Cuo, N. Voisin, et al. (2010), Implications of 21st century climate change for the hydrology of Washington State, Climatic Change, 102, 225, doi:10.1007/s10584-010-9855-0.

Evensen, G. (2007), Data assimilation: The ensemble Kalman filter, Springer, Berlin.

Gutmann, E., J. Hamman, and A. Wood (2017), NCAR/GARD: v0.4 (Version v0.4), Zenodo, doi:10.5281/zenodo.376874.

Hamill, T. M., and J. S. Whitaker (2006), Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application, Mon. Wea. Rev., doi:10.1175/MWR3237.1.

Hamill, T. M., J. S. Whitaker, and S. L. Mullen (2006), Reforecasts: An important dataset for improving weather predictions, Bull. Amer. Meteor. Soc., 87, 33–46, doi:10.1175/BAMS-87-1-33.

Hamill, T. M., G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau Jr., Y. Zhu, and W. Lapenta (2013), NOAA's second-generation global medium-range ensemble reforecast dataset, Bull. Am. Meteor. Soc., 94, 1553-1565, doi:10.1175/BAMS-D-12-00014.1.

Hartman, R. (2014), Operational data assimilation at NWS River Forecast Centers, presented at Catchment-based Hydrological Model Data Assimilation (CAHMDA VI) and Hydrologic Ensemble Prediction Experiment (HEPEX-DAFOH II) Join Workshop, 8-12 Sep. 2014,

Austin, TX, U.S.A., http://www.jsg.utexas.edu/ciess/files/session8_Robert_Hartman.pdf [last accessed 17 Nov. 2017].

Haughton, N., and Coauthors (2016), The plumbing of land surface models: Is poor performance a result of methodology or data quality?, J. Hydrometeor., 17, 1705-1723, doi:10.1175/JHM-D-15-0171.1.

Hersbach, H. (2000), Decomposition of the continuous ranked probability score for ensemble prediction systems, Wea. Forecast., 15, 559-570.

Kitagawa, G. (1996), Monte Carlo Filter and Smoother for non-Gaussian nonlinear state space models, J. Comput. Graphic. Stats., 5, 1-25.

Kong, A., J. S. Liu, and W. H. Wong (1994), Sequential imputations and Bayesian missing data problems, J. Am. Stat. Assoc., 89, 278-288.

Koutsoyiannis, D., H. Yao, and A. Georgakakos (2008), Medium-range flow prediction for the Nile: A comparison of stochastic and deterministic methods, Hydrol. Sci. J., 53, 142-164, doi:10.1623/hysj.53.1.142.

Laio, F. and S. Tamea (2007), Verification tools for probabilistic forecasts of continuous hydrological variables, Hydrol. Earth Syst. Sci., 11, 1267-1277, https://doi.org/10.5194/hess-11-1267-2007.

Leisenring, M. and H. Moradkhani (2010), Snow water equivalent estimation using Bayesian data assimilation methods, Stoch. Env. Res. Risk A., 25, 253–270, doi:10.1007/s00477-010-0445-5.

Lguensat, R., P. Tandeo, P. Ailliot, M. Pulido, and R. Fablet (2017), The analog data assimilation, Mon. Wea. Rev., 4093-4107, doi:10.1175/MWR-D-16-0441.s1.

Li, T., T. P. Sattar, Q. Han, and S. Sun (2013), Roughening methods to prevent sample impoverishment in particle PHD filter, FUSION2013, arXiv:1306.3875v1.

Li, T., S. Sun, T. P. Sattar, and J. M. Corchado (2014), Fight sample degeneracy and impoverishment in particle filters: A review of intelligent approaches, Expert Syst. with Appl., 41, 3944-3954.

Liu, J. S., and R. Chen (1995), Blind deconvolution via sequential imputations, J. Am. Stat. Assoc., 90, 567-576.

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

Lorenz, E. N. (1969), Atmospheric predictability as revealed by naturally occurring analogues, J. Atmos. Sci., 26, 636-646.

Magnusson, J., A. Winstral, A. S. Stordal, R. Essery, and T. Jonas (2017), Improving physically based snow simulations by assimilating snow depths using the particle filter, Water Resour. Res., 53, 1125-1143, doi:10.1002/2016WR019092.

Maneewongvatana, S., and D. M. Mount (1999), Analysis of approximate nearest neighbor searching with clustered point sets, Comput. Geom., arXiv:cs/9901013v1.

McCarthy, P.M., R. Sando, S. K. Sando, and D. M. Dutton (2016), Methods for estimating streamflow characteristics at ungaged sites in western Montana based on data through water year 2009: U.S. Geological Survey Scientific Investigations Report 2015-5019.

McEnery, J., J. Ingram, Q. Duan, T. Adams, and L. Anderson (2005), NOAA's Advanced Hydrologic Prediction Service: Building pathways for better science in water forecasting, Bull. Am. Meteor. Soc., 375-385, doi:10.1175/BAMS-86-3-375.

McLaughlin, D. (2002), An integrated approach to hydrologic data assimilation: Interpolation, smoothing, and filtering, Adv. Water Resour., 25, 1275-1286, doi:10.1016/S0309-1708(02)00055-6.

Mendoza, P. A., A. W. Wood, E. Clark, E. Rothwell, M. P. Clark, B. Nijssen, L. D. Brekke, and J. R. Arnold (2017), An intercomparison of approaches for improving predictability in operational seasonal streamflow forecasting, Hydrol. Earth Syst. Sci., doi:10.5194/hess-21-3915-2017.

Menne, M.J., I. Durre, R.S. Vose, B.E. Gleason, and T.G. Houston (2012a), An overview of the Global Historical Climatology Network-Daily Database, J. Atmos. Ocean. Tech., 29, 897-910, doi:10.1175/JTECH-D-11-00103.1 [last accessed 20 Sept. 2016].

Menne, M.J., I. Durre, B. Korzeniewski, S. McNeal, K. Thomas, X. Yin, S. Anthony, R. Ray, R.S. Vose, B. E. Gleason, and T.G. Houston (2012b), Global Historical Climatology Network - Daily (GHCN-Daily), Version 3, NOAA National Climatic Data Center, http://dx.doi.org/10.7289/V5D21VHZ [last accessed 20 Sept. 2016].

Montzka, C., H. Moradkhani, L. Weihermuller, M. Canty, H. J. Hendricks Franssen, and H. Vereecken (2011), Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter, J. Hydrol., 399, 410–421, doi:10.1016/j.jhydrol.2011.01.020.

Moradkhani, H., K.-L. Hsu, H. Gupta, and S. Sorooshian (2005), Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resour. Res., 41, W05012, doi:10.1029/2004WR003604.

Moradkhani, H., K. Hsu, Y. Hong, and S. Sorooshian (2006), Investigating the impact of remotely sensed precipitation and hydrologic model uncertainties on the ensemble streamflow forecasting, Geophys. Res. Lett., 33, L12401, doi:10.1029/2006GL026855.

Moradkhani, H., C. M. DeChant, S. Sorooshian (2012), Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method, Water Resour. Res., 48, doi:10.1029/2012WR012144.

Musso, C., N. Oudjane, F. LeGland, A. Doucet, J. F. G. de Freitas, N. J. Gordon (2001), Improving regularised particle filters, in Doucet et al., (eds.), Sequential Monte Carlo Methods in Practice, New York: Springer-Verlag.

Newman, A. J., M. P. Clark, J. Craig, B. Nijssen, A. Wood, E. Gutmann, N. Mizukami, L. Brekke, J. R. Arnold (2015), Gridded ensemble precipitation and temperature estimates for the contiguous United States, J. Hydrometeor., doi:10.1175/JHM-D-15-0026.1.

Raff, D., L. Brekke, K. Werner, A. Wood, and K. White (2013), Short-Term Water Management Decisions: User Needs for Improved Climate, Weather, and Hydrologic Information. Bureau of Reclamation, U.S. Army Corps of Engineers and National Oceanic and Atmospheric Administration, Technical Report CWTS-2013-1, http://www.ccawwg.us/docs/Short-Term_Water_Management_Decisions_Final_3_Jan_2013.pdf [last accessed 17 Nov. 2017].

Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resour. Res., 46, W05521, doi:10.1029/2009WR008328.

Salamon, P., and L. Feyen (2010), Disentangling uncertainties in distributed hydrological modeling using multiplicative error models and sequential data assimilation, Water Resour. Res., 46, W12501, doi:10.1029/2009WR009022.

Slivinski, L., E. Spiller, A. Apte, and B. Sandstede (2015), A hybrid particle-ensemble Kalman filter for Lagrangian data assimilation, Mon. Wea. Rev., doi:10.1175/MWR-D-14-00051.1.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson (2008), Obstacles to High-Dimensional Particle Filtering, Mon. Wea. Rev., 136, 4629–4640, https://doi.org/10.1175/2008MWR2529.1.

Sumioka, S. S., D. L. Kresch, and K. D. Kasnick (1998), Magnitude and frequency of floods in Washington: U.S. Geological Survey Water-Resources Investigations Report 97-4277, 91 p.

van Leeuwen, P. J. (2009), Particle filtering in geophysical systems, Mon. Wea. Rev., 4089-4114, doi:10.1175/2009MWR2835.1.

Wagener, T., and H. V. Gupta (2005), Model identification for hydrological forecasting under uncertainty, Stoch. Environ. Res. Risk Assess., 19, 378-387, doi:10.1007/s00477-005-0006-5.

Weerts, A. H., and G. Y. H. El Serafy (2006), Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resour. Res., 42, W09403, doi:10.1029/2005WR004093.

Welles, E., S. Sorooshian, G. Carter, and B. Olsen (2007), Hydrologic verification: A call for action and collaboration, Bull. Am. Meteo. Soc., 88, 503-511, doi:10.1175/BAMS-88-4-503.

Wood, A. W. (2012), Dynamical-statistical approaches for hydrologic ensemble prediction, in WIRADA (2012) Water Information Research and Development Alliance: Science Symposium Proceedings, Melbourne, Australia, 1–5 August 2011, http://www.clw.csiro.au/publications/waterforahealthycountry/wirada/WIRADA_Science_Symposium_Proceedings.pdf [last accessed 5 Dec. 2017].

Wood, A. W., and D. P. Lettenmaier (2008), An ensemble approach for attribution of hydrologic prediction uncertainty, Geophys. Res. Lett., 35, L14401, doi:10.1029/2008GL034648.

Wood, A. W., E. Clark, P. A. Mendoza, B., Nijssen, A. J. Newman, M. P. Clark, J. Arnold, and K. C. Nowak (2016), Assessing the viability of 'over-the-loop' real-time short-to-medium range ensemble streamflow forecasts, Abstract H41L-07 presented at 2016 Fall Meeting, AGU, San Francisco, Calif., 12-16 Dec.

Yan, H., and H. Moradkhani, (2016), Combined assimilation of streamflow and satellite soil moisture with the particle filter and geostatistical modeling, Adv. Water Resour., 94, 364-378, doi:10.1016/j.advwatres.2016.06.002.

Yao, H., and A. Georgakakos (2001), Assessment of Folsom Lake response to historical and potential future climate scenarios: 2. Reservoir management, J. Hydrol., 249, 176-196.

Yapo, P. O., H. V. Gupta, S. Sorooshian (1998), Multi-objective global optimization for hydrologic models, J. Hydrol., 204, 83-97, doi:10.1016/S0022-1694(97)00107-8.

Yin, S., and X. Zhu (2015), Intelligent particle filter and its application to fault detection of nonlinear system, IEEE Trans. Indust. Electron., 62, 3852-3861, doi:10.1109/TIE.2015.2399396.

Zhang, H., S. Qin, J. Ma, and H. You (2013), Using residual resampling and sensitivity analysis to improve particle filter data assimilation accuracy, IEEE Geosci. Remote Sens. Lett., 27, 1404-1408, doi:10.1109/LGRS.2013.2258888.

Zhu, M., P. J. van Leeuwen, J. Amezcua (2016), Implicit equal-weights particle filter, Quart. J. Royal Meteor. Soc., 142, 1904-1919, doi:10.1002/qj.2784.

APPENDIX A

This appendix includes three supporting figures for Chapter 2. Figure A.1 shows the workflow used to resample particles when the number of effective particles falls below a threshold, as described in Section 2.2.2. Figures A.2 and A.3 include the daily and long-term average monthly streamflow values for the calibration period 1981-2008 at Howard Hanson and Hungry Horse, respectively. The observations, calibration, and simulations are described in section 2.3.2.

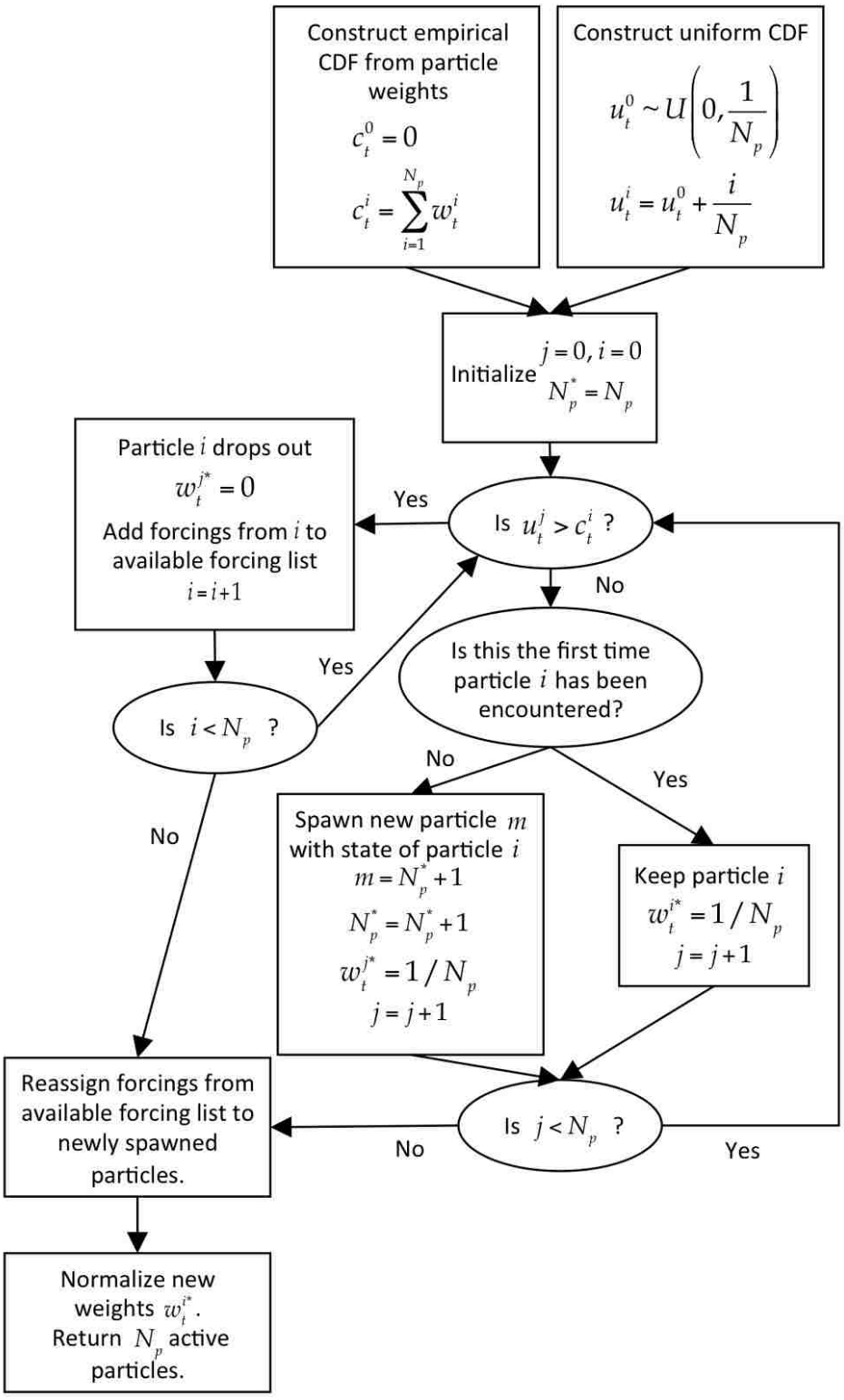

Figure A. 1. Resampling flow chart. See section 2.2 in the main text for detail.

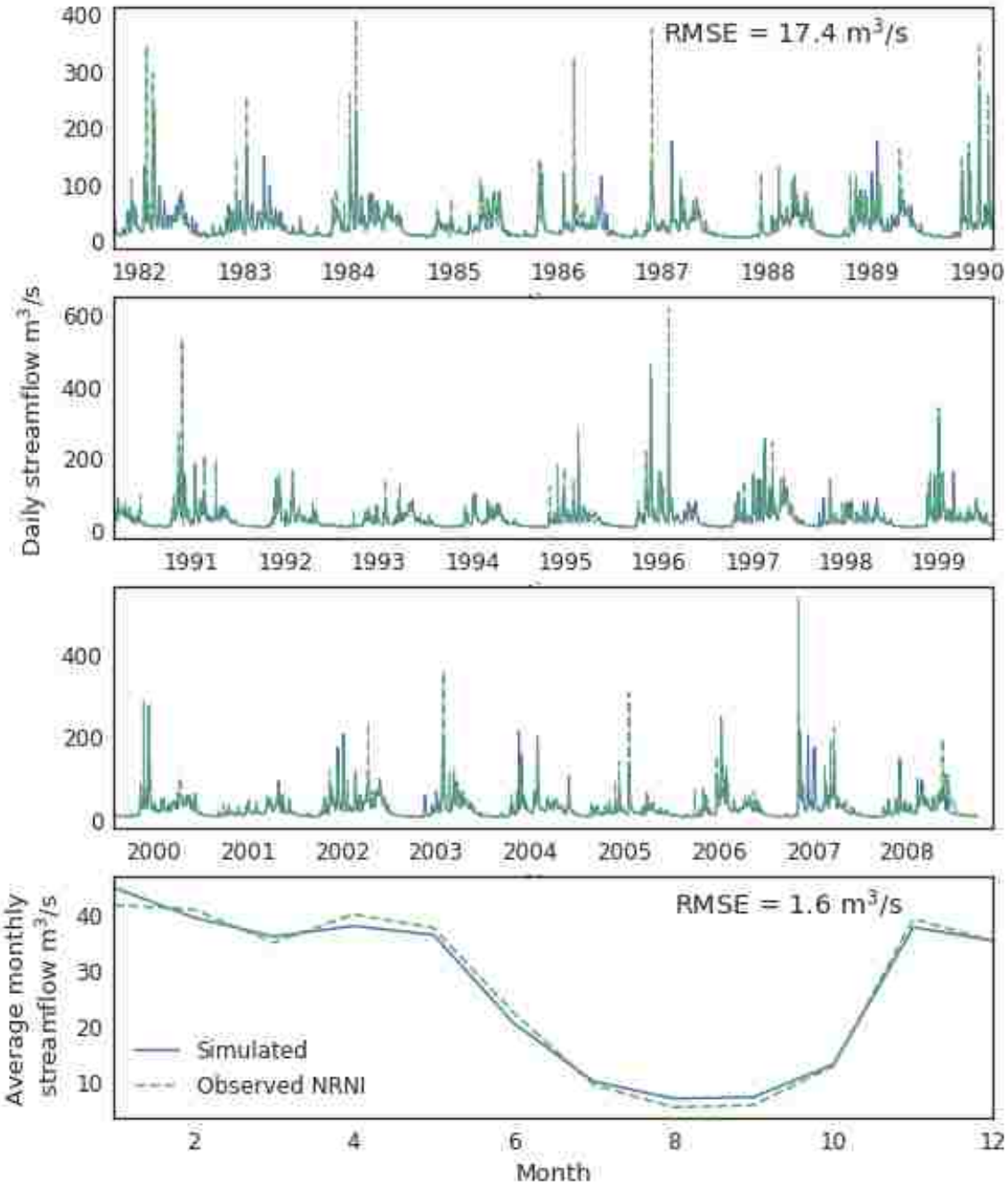

Figure A. 2. Daily streamflow (upper 3 panels) and long-term average monthly streamflow (lower panel) at Howard Hanson for calibration period from 1 October 1981 to 30 September 2008. The RMSE of daily flow is 17.3 m3/s and of long-term average monthly streamflow is 1.6 m^3/s.

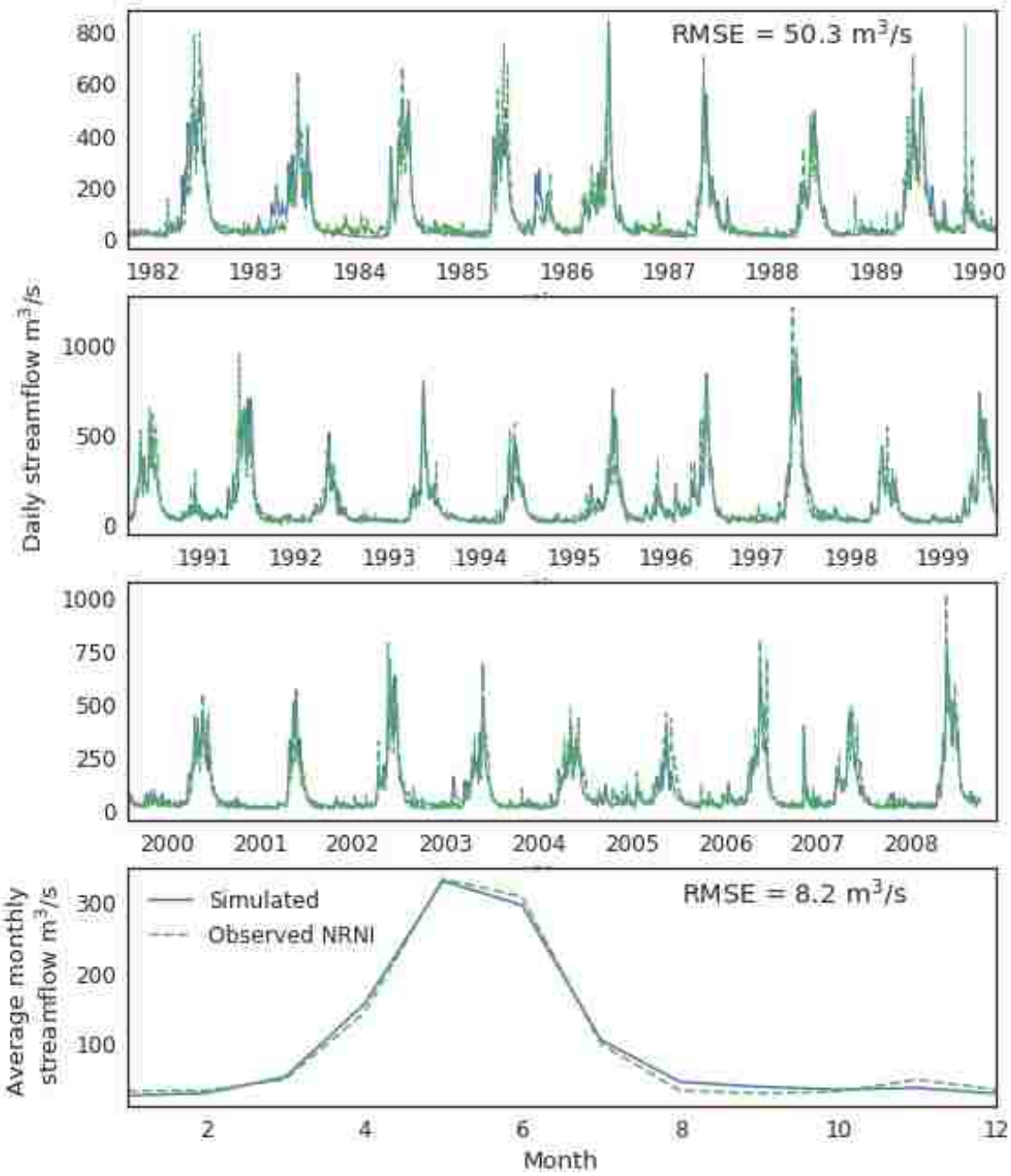

Figure A. 3. Daily streamflow (upper 3 panels) and long-term average monthly streamflow (lower panel) at Hungry Horse for calibration period from 1 October 1981 to 30 September 2008. The RMSE of daily flow is 50.3 m^3/s and of long-term average monthly streamflow is 8.2 m^3/s.

VITA

EDUCATION

Ph.D., 2017 Civil and Environmental Engineering, University of Washington, Seattle, WA
Thesis topic: *Particle Filter data assimilation of streamflow in basins with seasonal snow for initializing short- to medium-range streamflow forecasts* (Advisor: Bart Nijssen)

M.S.E., 2006 Civil and Environmental Engineering, University of Washington, Seattle, WA
Thesis: *Prospects for river discharge estimation through assimilation of remotely-sensed altimetry into a raster-based hydraulics model* (Advisor: Dennis Lettenmaier),

B.A., 2002 Magna cum laude, Geology, Carleton College, Northfield, MN
Thesis: *Analysis of the deformational and metamorphic history of cordierite schist, East Gulch, Colorado*

PUBLICATIONS IN PRESS

Clark, E. A., S. Biancamaria, F. Hossain, J.-F. Crétaux, and D.P. Lettenmaier (2014), Altimetry applications to transboundary river management, Chapter 13, in J. Benveniste, S. Vignudelli, and A. Kostianoy (Eds.), *Inland Water Altimetry*, Springer, accepted.

PEER-REVIEWED PUBLICATIONS

Adam, J. C., **E. A. Clark**, D. P. Lettenmaier, and E. F. Wood (2006), Correction of global precipitation products for orographic effects, J. Clim., 19, 15-38.

Andreadis, K. M., **E. A. Clark**, A. W. Wood, A. F. Hamlet, and D. P. Lettenmaier (2005), Twentieth Century drought in the conterminous United States, J, Hydromet., 6, 985-1001.

Andreadis, K. M., **E. A. Clark**, D. P. Lettenmaier, and D. E. Alsdorf, (2007), Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model, Geophys. Res. Lett., 34, L10403, doi:10.1029/2007GL029721.

Biancamaria, S., M. Durand, K. M. Andreadis, P. D. Bates, A. Boone, N. M. Mognard, E. Rodríguez, D. E. Alsdorf, D. P. Lettenmaier, and **E. A. Clark** (2011), Assimilation of virtual wide swath altimetry to improve Arctic river modeling, Remote Sens. Environ., 115, 373 - 381.

Biancamaria, S., K. M. Andreadis, M. Durand, **E. A. Clark**, E. Rodriguez, N. Mognard, D. Alsdorf, D. P. Lettenmaier, and Y. Oudin (2010), Preliminary characterization of SWOT hydrology error budget and global capabilities, J. Select. Topics Earth Obs. Remote Sens., 3, 6-19, doi: 10.1109/JSTARS.2009.2034614.

Clark, E. A., J. Sheffield, M. T. H. van Vliet, B. Nijssen, and D. P. Lettenmaier (2014), Continental runoff into the oceans (1950-2008), J. Hydromet., 16, 1502-1520, doi:10.1175/JHM-D-14-0183.1.

Groisman, P. Ya., **E. A. Clark**, V. M. Kattsov, D. P. Lettenmaier, I.N. Sokolik, V. B. Aizen, O. Cartus, J. Chen, S. Conard, J. Katzenberger, O. Krankina, J. Kukkonen, T. Machida, S. Maksyutov, D. Ojima, J. Qi, V. E. Romanovsky, M. Santoro, C. C. Schmullius, A. I. Shiklomanov, K. Shimoyama, H. H. Shugart, J. K. Shuman, M. A. Sofiev, A. I. Sukhinin, C. Vorosmarty, D. Walker, and E. F. Wood (2009), The Northern Eurasia Earth Science Partnership: An example of science applied to societal needs, Bull. Amer. Meteor. Soc., 90, 671-688, doi:10.1175/2008BAMS2556.1.

Lanini, J. L., **E. A. Clark**, and D. P. Lettenmaier (2009), Effects of fire-precipitation timing and regime on post-fire sediment delivery in Pacific Northwest forests, Geophys. Res. Lett., 36, L01402, doi:10.1029/2008GL034588.

Mendoza, P. A., A. W. Wood, **E. Clark**, E. Rothwell, M. P. Clark, B. Nijssen, L. D. Brekke, and J. R. Arnold (2017), An intercomparison of approaches for improving predictability in operational seasonal streamflow forecasting, Hydrol. Earth Syst. Sci., doi:10.5194/hess-21-3915-2017.

L'Ecuyer, T., H. K. Beaudoing, M. Rodell, W. Olson, B. Lin, S. Kato, C. A. Clayson, E. Wood, J. Sheffield, R. Adler, G. Huffman, M. Bosilovich, G. Gu, F. Roberston, P. Houser, D. Chambers, J. Famiglietti, E. Fetzer, W. T. Liu, X. Gao, C. A. Schlosser, **E. Clark**, D. P. Lettenmaier, and K. Hilburn (2015), The observed state of the energy budget in the early 21st Century, J. Clim., doi:10.1175/JCLI-D-14-00556.1.

Mehran, A., **E. Clark**, and D. Lettenmaier (2017), Spatial variability of wet troposphere delays over inland water bodies, J. Geophys. Res.-Atmos., doi: 10.1002/2017JD026525.

Rodell, M., H. K. Beaudoing, T. S. L'Ecuyer, W. S. Olson, J. S. Famiglietti, P. R. Houser, R. Adler, M. G. Bosilovich, C. A. Clayson, D. Chambers, **E. Clark**, E. J. Fetzer, X. Gao, G. Gu, K. Hilburn, G. J. Huffman, D. P. Lettenmaier, W. T. Liu, C. A. Schlosser, J. Sheffield, and E. F. Wood (2015), The observed state of the water cycle in the early 21st Century, J. Clim., doi:10.1175/JCLI-D-14-00555.1.

Rosenberg, E. A., **E. A. Clark**, A.C. Steinemann, and D. P. Lettenmaier (2013), On the contribution of groundwater storage to interannual streamflow anomalies in the Colorado River basin, Hydrol. Earth Syst. Sci., 17, 1475-1491, doi:10.5194/hess-17-1475-2013.

Yoon, Y., M. Durand, C. J. Merry, **E. A. Clark**, K. M. Andreadis, and D. E. Alsdorf (2012), Estimating river bathymetry from data assimilation of synthetic SWOT measurements, J. Hydrol., 464-465, 363-375, doi:10.1016/j.jhydrol.2012.07.028.

HONORS AND AWARDS

NASA Graduate Student Education Fellowship, *Jet Propulsion Laboratory*	2012-14
NASA Graduate Student Research Program Fellowship, *Jet Propulsion Laboratory*	2011-12
Stanford Graduate Fellowship, *Stanford University*	2005-08
Andy Studebaker Fellowship, *Univ. Washington, Center for Water & Watershed Studies*	2004
Duncan Stewart Fellowship, *Carleton College, Department of Geology*	2001-02

REGISTRATION OR PROFESSIONAL CERTIFICATION

Engineer-In-Training (EIT), Washington State, License No. 27833