

# Daily Data Assimilation of a Hydrologic Model Using the Ensemble Kalman Filter

*Sami Andrew Malek*



Electrical Engineering and Computer Sciences  
University of California at Berkeley

Technical Report No. UCB/EECS-2019-101

<http://www2.eecs.berkeley.edu/Pubs/TechRpts/2019/EECS-2019-101.html>

June 14, 2019

Copyright © 2019, by the author(s).  
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

### Acknowledgement

This work was partially supported by the Civil Systems department at UC Berkeley, California Energy Commission through the grant “Improving Hydrological Snow pack Forecasting for Hydropower Generation Using Intelligent Information Systems” (EPC-14-067), Pacific Gas and Electric Co (PG&E), the California Department of Water Resources and UC Water.

I would like to thank Prof. Glaser for recommending me to pursue this opportunity and to thank Prof. Bayen for his advice and words of encouragement.

Finally, special thanks to Kevin Richards (PG&E) for closely collaborating by sharing streamflow measurements and Feather river data pertaining to PRMS such as HRUs and calibrated parameters.

**Daily Data Assimilation of a Hydrologic Model Using the Ensemble Kalman  
Filter**

by

Sami A. Malek

B.E. (American University of Beirut) 2013  
M.S. (University of California at Berkeley) 2014

A project report submitted in partial satisfaction  
of the requirements for the degree of

Master of Science

in

Engineering - Electrical Engineering and Computer Science

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Alexandre M. Bayen, Chair  
Professor Steven D. Glaser

Summer 2019

---

**Daily Data Assimilation of a Hydrologic Model Using the  
Ensemble Kalman Filter**

by Author Sami A. Malek

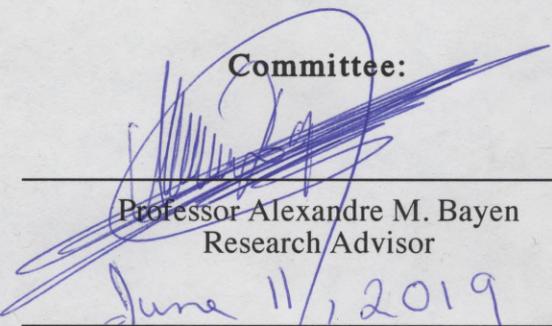
---

**Research Project**

Submitted to the Department of Electrical Engineering and Computer Sciences,  
University of California at Berkeley, in partial satisfaction of the requirements for the  
degree of **Master of Science, Plan II.**

Approval for the Report and Comprehensive Examination:

**Committee:**



---

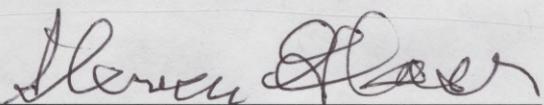
Professor Alexandre M. Bayen  
Research Advisor

June 11, 2019

---

Date:

\*\*\*\*\*



---

Professor Steven D. Glaser  
Second Reader

6-13-2019

---

Date:

Daily Data Assimilation of a Hydrologic Model Using the Ensemble Kalman Filter

Copyright © 2019

by

Sami A. Malek

## Abstract

Daily Data Assimilation of a Hydrologic Model Using the Ensemble Kalman Filter

by

Sami A. Malek

Master of Science in Engineering - Electrical Engineering and Computer Science

University of California, Berkeley

Professor Alexandre M. Bayen, Chair

Accurate runoff forecasting is crucial for reservoir operators as it allows optimized water management, flood control and hydropower generation. Land surface models in mountainous regions depend on climatic inputs such as precipitation, temperature and solar radiation to model the water and energy dynamics and produce runoff as output. With the rapid development of cheap electronics applied in various systems, such as Wireless Sensor Networks (WSNs), satellite and airborne technologies, the prospect of practically measuring spatial Snow Water Equivalent in a dense temporal scale is increasing. We present a framework for updating the Precipitation Runoff Modeling System (PRMS) with Snow Water Equivalent (SWE) maps and runoff measurements on a daily timescale based on the Ensemble Kalman Filter (ENKF). Results show that by assimilating SWE daily, the modeled SWE gets updated accordingly, however no improvement is observed at the runoff model output. Instead, a deterioration consistently occurs. Augmenting the state space with model parameters and runoff model output allows for filter update with previous day measured runoff using the joint state-parameter method, and showed a considerable improvement in the daily runoff output of up to 60% reduction in RMSE for the wet water year 2011 relative to the no assimilation scenario, and improvement of up to 28% compared to a naive autoregressive AR(1) filter. Additional simulation years showed consistent improvement compared to no

assimilation, but varied relative to the previous day autoregressive forecast during the dry year 2014.

---

Professor Alexandre M. Bayen  
Masters Project Report Chair

I dedicate this work to my father and my advisors.

# Contents

<b>Contents</b>	<b>ii</b>
<b>List of Figures</b>	<b>iv</b>
<b>List of Tables</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vi</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 System Dynamics</b>	<b>3</b>
2.1 Precipitation Runoff Modeling System . . . . .	3
2.1.1 Model Inputs . . . . .	7
2.1.2 Interception . . . . .	9
2.1.3 Evapotranspiration . . . . .	10
2.1.4 Snowpack . . . . .	10
2.1.5 Impervious Zone Reservoir . . . . .	12
2.1.6 Soil Zone Reservoir . . . . .	13
2.1.7 Surface Runoff . . . . .	13
2.1.8 Subsurface Reservoir . . . . .	14
2.1.9 Ground-Water Reservoir . . . . .	14
2.1.10 Model Output: Streamflow . . . . .	15
2.2 Model Features . . . . .	15
<b>3 Filtering Techniques</b>	<b>17</b>
3.1 Overview . . . . .	17
3.2 EnKF & State-space . . . . .	21

<b>4</b>	<b>Experimental Results</b>	<b>27</b>
4.1	SWE State Assimilation . . . . .	28
4.2	SWE and Runoff Joint State-Parameter Assimilation . . . . .	30
4.2.1	New State Space . . . . .	30
4.2.2	Results . . . . .	32
<b>5</b>	<b>Conclusion</b>	<b>35</b>
	<b>Bibliography</b>	<b>37</b>
	References . . . . .	37
<b>A</b>	<b>Model Reference</b>	<b>40</b>
A.1	Model Parameters Reference . . . . .	40
A.2	Model Variables Reference . . . . .	40
<b>B</b>	<b>Additional Simulation Years</b>	<b>43</b>
B.1	Water Year 2011 Results . . . . .	43
B.1.1	SWE update . . . . .	43
B.1.2	Joint update . . . . .	43
B.2	Water Year 2014 Results . . . . .	43
B.2.1	SWE update . . . . .	43
B.2.2	Joint update . . . . .	43

# List of Figures

2.1	Hydrologic Response Units (HRUs) of the East Branch of the Feather River overlaid on a SWE map from 1998. . . . .	4
2.2	Schematic diagram of PRMS and its inputs [1] . . . . .	5
4.1	Water year 2006 basin streamflow output. . . . .	29
4.2	Water year 2006 basin average SWE. . . . .	33
4.3	Water year 2006 basin average streamflow. . . . .	34
B.1	Water year 2011 basin average streamflow (SWE update). . . . .	44
B.2	Water year 2011 basin average SWE. (SWE or Joint update) . . . . .	45
B.3	Water year 2011 basin average streamflow (Joint update). . . . .	46
B.4	Water year 2014 basin average streamflow (SWE update). . . . .	47
B.5	Water year 2014 basin average SWE. (SWE or Joint update) . . . . .	48
B.6	Water year 2014 basin average streamflow. (Joint update) . . . . .	49

# List of Tables

2.1	Model Inputs . . . . .	7
2.2	Model States . . . . .	7
4.1	Perturbed and updated parameters. . . . .	31
4.2	Experiments performance in terms of streamflow output RMSE . . . . .	32
4.3	Experiments validation in terms of basin-mean SWE RMSE. . . . .	32
A.1	Model parameters . . . . .	41
A.2	Model variables . . . . .	42

## Acknowledgements

This work was partially supported by the Civil Systems department at UC Berkeley, California Energy Commission through the grant Improving Hydrological Snow pack Forecasting for Hydropower Generation Using Intelligent Information Systems (EPC-14-067), Pacific Gas and Electric Co (PG&E), the California Department of Water Resources and UC Water.

I would like to thank Prof. Glaser for recommending me to pursue this opportunity and to thank Prof. Bayen for his advice and words of encouragement.

Finally, special thanks to Kevin Richards (PG&E) for closely collaborating by sharing streamflow measurements and Feather river data pertaining to PRMS such as HRUs and calibrated parameters.

# Chapter 1

## Introduction

This research work fits into a larger objective of attaining consistently more accurate snowmelt runoff forecast resilient to climate change and extreme weather conditions. Presently, reservoir and dam operators rely on forecasts largely based on historical methods. Numerous studies show that climate change effects are challenging the stationarity assumption of this statistical approach. The past can no longer predict the future as well as it did before. Other studies show a systematic receding date for snow melt during the year. [2]–[5]

A mandatory evacuation order was issued in February 2017 after the Oroville dam crisis because of hazardous water overflow at the Oroville reservoir. The Department of Water Resources (DWR) published a report showing that reservoir water levels rose earlier and higher than previous 16 years. Oroville is the pouring point of the North Fork Feather River where 3 days earlier, recently deployed sensors showed an onset of rapid snow melt up to 1 cm per hour most probably due to a rain on snow event. Unlike statistical models, a physically based forecast model that incorporates these new data should be able to capture such inflow surges.

In addition to flood control, runoff forecasting allows for better reservoir management and potentially more efficient power generation. Stakeholders, such as Pacific Gas & Electric (PG&E) have already started shifting towards physically based forecast models. These

models though less vulnerable to climate change, require dense and accurate data as well as a methodology to seamlessly integrate sensor measurements of different temporal and spatial scales.

Recent efforts to increase hydrologic instrumentation in the East Branch of the North Fork Feather River basin by a partnership between the University of California at Berkeley, the Department of Water Resources and the California Energy Commission (CEC) led to the deployment of four 1 km<sup>2</sup> scale state-of-the-art remote autonomous ground measurement systems based on Wireless Sensor Networks technology to collect temperature, humidity, solar radiation, snowdepth, soil temperature and soil moisture measurements at a 15 minute scale. These new deployments would supplement the existing measurement system consisting of 8 snow pillows reporting daily SWE measurements and the monthly manual snow course locations. Moreover the NASA airborne snow observatory, in partnership with DWR are using airborne LiDAR system to produce unprecedented bi-weekly basin-wide SWE maps of the Tuolumne and Merced river basins at 50 meter spatial resolution [6].

The objective of this study is to design a practical data assimilation framework for real-time runoff forecasting using PRMS that can leverage sensor data from the existing, state-of-the-art and future hydrologic measurement systems at different temporal and spatial scales.

After describing the PRMS model dynamics, we present existing filtering techniques, then we develop a practical data assimilation framework to integrate in real-time simulated measurements of SWE and real measurements of streamflow in the model using the Ensemble Kalman Filter.

## Chapter 2

# System Dynamics

### 2.1 Precipitation Runoff Modeling System

The Precipitation Runoff Modeling System is a hydrologic model originally described by Leavesley (1983). It simulates snow accumulation and ablation processes as well as soil-zone and streamflow water dynamics. The watershed or basin is discretized into Hydrologic Response Units (HRUs) based on topography, land use, climate, soil properties, and geologic units information [7]. Sample HRUs for the East Branch of the Feather River basin are shown in Fig. 2.1.

The model runs on a daily time discretization, with some processes having half-day resolution. The model takes daily precipitation, minimum and maximum air temperature and optionally short-wave solar radiation as inputs. PRMS is a module based system: it can handle other data inputs as well as different modeling modules. We restrict this study to what is currently being operationally used in our specific application. Each HRU consists of a system of abstract water containers (or storages) interconnected by water flows as shown in Fig. 2.2. Variables with subscripts “h” in this report indicate that those variables are unique for each HRU.

The components are Interception ( $Sint_h$ ), Snowpack ( $SW E_h$  and  $H_h$ ), Impervious Zone Reservoir ( $Simp_h$ ), Soil Zone Reservoir ( $Ssz_h$ , subdivided into Recharge  $Ssre_h$  and Lower

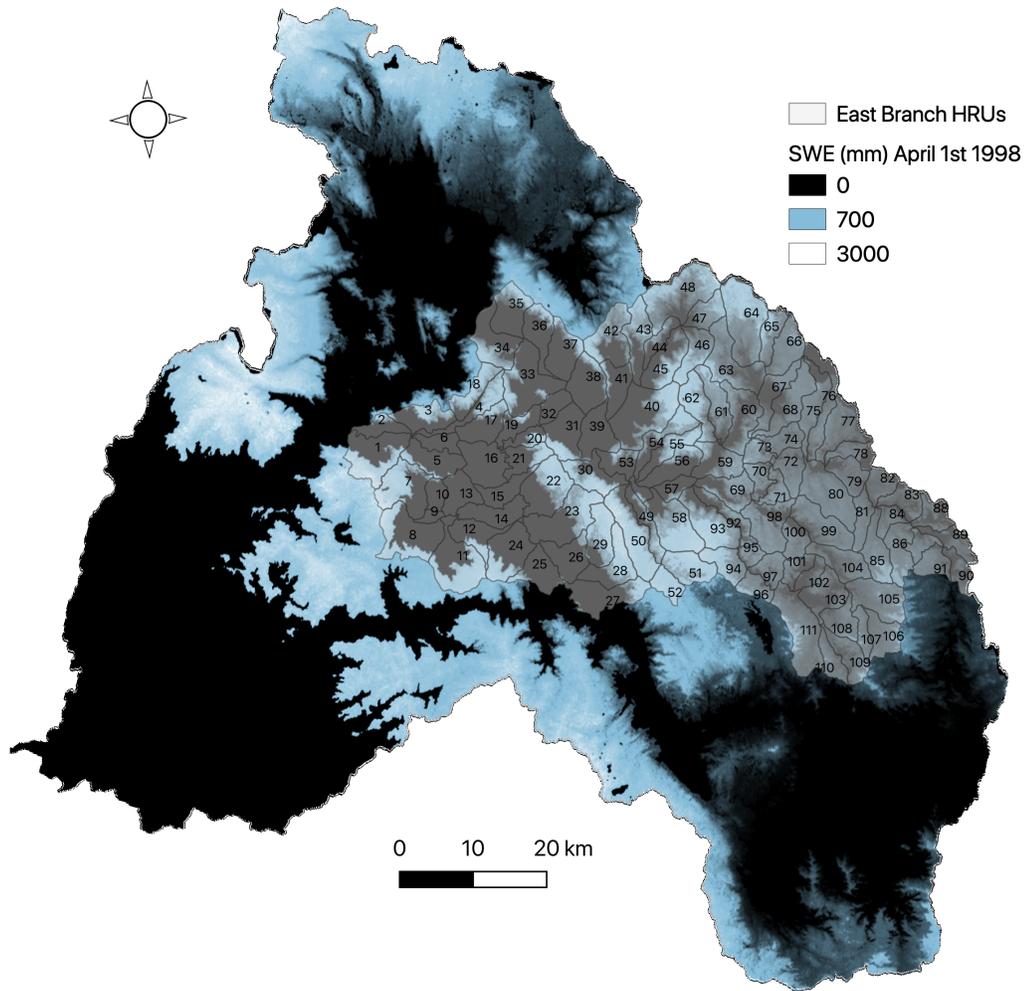


Figure 2.1: Hydrologic Response Units (HRUs) of the East Branch of the Feather River overlaid on a SWE map from 1998.

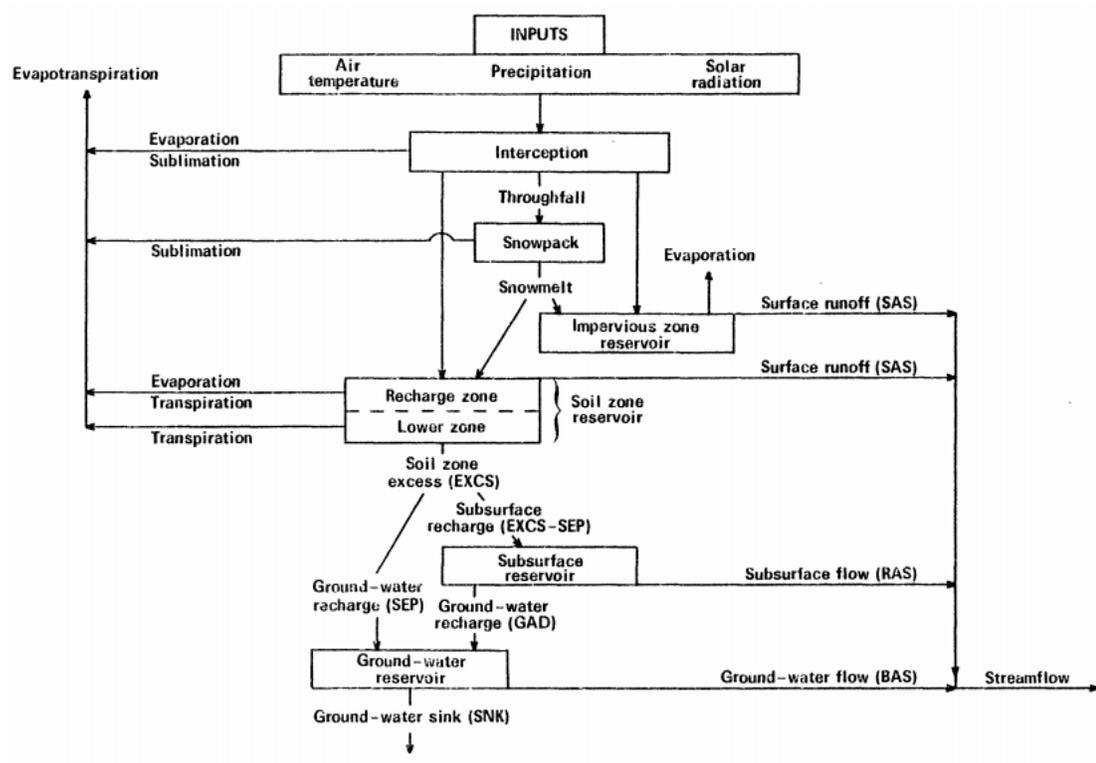


Figure 2.2: Schematic diagram of PRMS and its inputs [1]

zone  $Ssl_h$ ), Subsurface Reservoir ( $Sss_h$ ) and the Ground-Water Reservoir ( $Sgw_h$ ). For each water storage, the water mass balance equation is satisfied.

$$\frac{\Delta(S)}{\Delta t} = \sum F \quad (2.1)$$

where  $S$  and  $F$  denote water storage and flow respectively.

$$\frac{\Delta(Sint)}{\Delta t} = \sum Fint_{in} - \sum Fint_{out} \quad (2.2)$$

where  $Fint_{in}$  consists of precipitation, and  $Fint_{out}$  consists of evaporation, sublimation and throughfall.

$$\frac{\Delta(SWE)}{\Delta t} = \sum Fsw_{e_{in}} - \sum Fsw_{e_{out}} \quad (2.3)$$

where  $Fsw_{e_{in}}$  consists of precipitation on non-vegetated areas, throughfall from canopy covered regions, and  $Fsw_{e_{out}}$  consists of sublimation and snowmelt.

$$\frac{\Delta(Ssz)}{\Delta t} = \sum Fsz_{in} - \sum Fsz_{out} \quad (2.4)$$

where  $Fsz_{in}$  consists of snowmelt, rain, and flow from interception on pervious regions, and  $Fsz_{out}$  consists of surface runoff, evapotranspiration and excess  $Sz$  water.

$$\frac{\Delta(Sss)}{\Delta t} = \sum Fss_{in} - \sum Fss_{out} \quad (2.5)$$

where  $Fss_{in}$  consists of recharge from  $Ssz$ , subsurface flow to streamflow, and percolation to the ground water reservoirs ( $Sgw$ ).

$$\frac{\Delta(Sgw)}{\Delta t} = \sum Fsgw_{in} - \sum Fsgw_{out} \quad (2.6)$$

where  $Fsgw_{in}$  consists of recharge from Soil Zone and from subsurface reservoirs and  $Fsgw_{out}$  consists of ground water sink and groundwater flow to streamflow output.

For the Snowpack storage( $SWE$ ), melt occurs when the snowpack reaches isothermal conditions ( $H = 0$ ). An additional energy balance equation is satisfied that deals with

the atmospheric and temperature gradient and energy fluxes ( $f$ ) interactions and can be written as:

$$\frac{\Delta(H)}{\Delta t} = \sum f \quad (2.7)$$

Symbol	PRMS symbol	Type	Quantity	Units	Description
P	hru_ppt	daily	per HRU	inches	total precipitation
Tmax, i	tmax	daily	one	°F	max temperature at HRU i
Tmin, i	tmin	daily	one	°F	min temperature at HRU i

Table 2.1: Model Inputs

Symbol	PRMS symbol	Quantity	Units	Description
Sint	intcp_stor	per HRU	inches	interception storage
Sliq	freeh2o	per HRU	inches	liquid water snowpack storage
	pk_den	per HRU	g/cm <sup>3</sup>	snowpack average density
SWE	pkwater_equiv	per HRU	inches	total water snowpack storage
Sice	pk_ice	per HRU	inches	frozen water snowpack storage
D	pk_depth	per HRU	inches	snowpack depth
H	pk_def	per HRU	Langleys	snowpack heat deficit
Tpk	pk_temp	per HRU	°F	snowpack temperature
fzca	snowcov_area	per HRU	-	snow cover area
Sgw	gwres_stor	per GWR	inches	ground water storage
Simp	imperv_stor	per HRU	inches	impervious storage
Ssre	soil_rechr	per HRU	inches	soil recharge storage
Ssz	soil_moist	per HRU	inches	total soil-zone storage (soil-moisture)
Sss	ssres_stor	per SSR	inches	sub-surface storage
SWEmax	pst	per HRU	inches	tracks max SWE of pack for fzca albedo

Table 2.2: Model States

### 2.1.1 Model Inputs

Model inputs consist of daily time-series of precipitation for each HRU and minimum and maximum air temperature for 1 HRU that contains a temperature station. The daily short-wave radiation ( $R_{sw}$ ) is estimated internally by the model. They are detailed below:

Short-Wave Radiation: A constant solar table consisting of daily estimates of the potential (clear sky) short-wave solar radiation ( $R_{psw}$ ) for each HRU (they are derived from calculations of duration of sun exposure and latitude for each HRU and each day [1]). Such

values are then adjusted daily based on a modification of the “degree-day” method to get the short-wave radiation described in more details in [1]. The method consists of deriving a degree-day coefficient ( $dd$ ) and then a ratio of actual-to-potential radiation for a horizontal surface ( $r$ ). The following parametric linear curve is used to find  $dd$ :

$$dd = \phi_m \cdot Tmax_f + \beta_m \quad (2.8)$$

where  $\phi_m$  and  $\beta_m$  are parameters and  $Tmax_f$  is an input. From  $dd$ ,  $r_h$  is retrieved via a non-linear relationship of  $dd$  in function of  $r_h$ . Finally the actual solar radiation is adjusted according to each HRU slope.

$$Rsw_h = \frac{r_h \times \gamma}{\cos(\arctan(\delta_h))} \cdot Rpsw_h \quad (2.9)$$

where

$$\gamma = \begin{cases} \gamma_s & \text{during summer days \& } P \text{ greater than } \mathbf{Pmin} \\ \gamma_w & \text{during winter days \& } P \text{ greater than } \mathbf{Pmin} \\ 1 & P \text{ less than } \mathbf{Pmin} \end{cases} \quad (2.10)$$

because  $Rpsw_h$  is calculated for days without precipitation.

### Air Temperature

The single temperature input at HRU  $i$  with elevation  $e_i$  is spatially distributed for the remaining HRUs according to the following equation (and similarly for  $T_{min}$ ) in degrees Fahrenheit.

$$Tmax_h = Tmax_{h_i} - \lambda_{1month} \times \left( \frac{e_h - e_{h_i}}{1000} \right) - \beta_{1h} \quad (2.11)$$

$$Tmin_h = Tmin_{h_i} - \lambda_{2month} \times \left( \frac{e_h - e_{h_i}}{1000} \right) - \beta_{2h} \quad (2.12)$$

Precipitation Precipitation  $P_h$  for each HRU are inputs to the model. The precipitation phase is determined solely based on temperature. If the HRU maximum air temperature ( $Tmax_h$ ) is less than or equal to the parameter  $Ts$ , all precipitation is snow. If  $Tmin_h$  and  $Tmax_h$  are greater than or equal to  $Ts$  and  $Tr$  respectively, precipitation is assumed to be

all rain. Otherwise,  $P$  is considered a mixture of rain and snow. The rain fraction of  $P$  is computed as:

$$fr_h = \left( \frac{Tmax_h - Ts}{Tmax_h - Tmin_h} \right) \cdot \zeta m \quad (2.13)$$

A value greater than 1 is considered an all rain event.

### 2.1.2 Interception

Portions of the HRU with canopy cover intercept precipitation. Different canopy cover density values are used for the summer ( $ds_h$ ) and winter ( $dw_h$ ). Only when the precipitation amount exceeds the canopy storage capacity ( $Sc$ ) does it reach the ground as throughfall ( $Pt_h$ ). For example the net precipitation reaching the ground during summer is:

$$Pn_h = Pt_h \cdot ds_h + (1.0 - ds_h) \times P_h \quad (2.14)$$

where the throughfall  $Pt_h$  is computed as:

$$Pt_h = \begin{cases} P_h - \frac{Sc_h}{A_h \cdot ds_h} & : P_h > \frac{Sc_h}{A_h \cdot ds_h} \\ 0.0 & : \text{otherwise} \end{cases} \quad (2.15)$$

where the canopy storage capacity  $Sc_h$  is:

$$Sc_h = (Crs_h - Sint_h) \cdot A_h \cdot ds_h \quad (2.16)$$

Parameters  $ds_h$  and  $Crw_h$  are used during winter. A similar process is modeled for snow.  $Sint_h$  is a model state holding intercepted precipitation.  $Crs_h$  and  $Crw_h$  are the summer and winter rain interception storage capacity for the major vegetation type for each HRU.  $Cs_h$  is used for snow. Intercepted rain, (or snow) evaporates (or sublimates) at a free-water surface rate.

### 2.1.3 Evapotranspiration

The Jensen-Haise formulation [1] is used to compute the potential evapotranspiration:

$$ET_h = \mathbf{j}c \cdot (T_h - jc_h) \cdot \frac{Rsw_h}{2.54 \cdot (597.3 - (0.5653 \cdot T_h))} \quad (2.17)$$

where the coefficient  $jc_h$  is approximated to be:

$$jc_h = 22 - \frac{e_h}{1000} \quad (2.18)$$

As figure 2.2 shows, evapotranspiration occurs in multiple reservoirs.

### 2.1.4 Snowpack

Snowpack dynamics are modeled using water and energy balances for each HRU. The snowpack is abstracted as 2 layers. The energy exchanges that occur between the snowpack and the atmosphere are radiative, conductive and convective. The energy balance at the snow-atmosphere interface is computed twice per day, for both the day and night periods as:

$$\frac{dE}{dt} = I_h + Qv_h + R_h \quad (2.19)$$

$E$  is the surface energy of the snowpack

The net short wave radiation  $R_h$  in (2.19) is computed from  $Rsw_h$  (2.9) after accounting for the reflected portion of the radiation using the surface albedo  $\alpha$  and that limited by the vegetative transmission coefficient parameter  $\psi_h$ .

$$R_h = (1 - \alpha_h) \cdot \psi_h \cdot Rsw_h \quad (2.20)$$

Where  $\alpha$  is obtained from a non-linear curve that is a function of the snow dynamics phase and time since last new snow in days.

The net incoming longwave radiation  $I_h$  originates from the atmosphere and land cover:

$$I_h = \mathbf{c}w_h \cdot Ip_h + (1 - \mathbf{c}w_h) \cdot \epsilon \cdot Ip_h \quad (2.21)$$

where  $\epsilon$  is the emissivity, a precipitation situation-dependent parameter (between 0.0 and 1.0 dependent on storm type), and  $I_{ph}$  is the perfect black-body emission:

$$I_{ph} = 5.85 \cdot 10^{-8} \cdot (T_h + 273.16)^4 \quad (2.22)$$

The combined convection and latent heat flux from condensation is modeled as linear function of temperature.

$$Qv_h = \omega \cdot T_h \quad (2.23)$$

When (2.19) is negative, heat flow occurs by conduction only (ex: potentially refreeze snow at isothermal) and depends on the snowpack density and layer temperature gradient:

$$Qc_h = 2 \cdot \left[ 0.5 \cdot \sigma \sqrt{\frac{0.0077 \cdot \rho_h^2 \cdot \Delta t}{\sigma \cdot \rho_h}} \right] \cdot (T_h - Tpk_h) \quad (2.24)$$

where  $\sigma$  is the specific heat of ice (in Celsius  $Cal.m^{-3} \cdot ^\circ C^{-1}$ ). A positive energy exchange translates into snowmelt at the surface of the snowpack, that transports heat  $\Delta E$  into the lower snowpack by mass transfer. The potential melt from such heat is:

$$M_h = \frac{\Delta E}{203.2} \cdot sca_h \quad (2.25)$$

where  $sca_h = f sca_h \cdot A_h$  is the HRU snow covered area.

The energy state of the lower layer is represented as a heat deficit. The snowpack heat deficit ( $H$ ) for each HRU represents the amount of heat necessary to bring the snowpack to isothermal 0 °C. Snowmelt occurs only when the heat deficit reaches zero and the snowpack free-water storage is exceeded. Otherwise, potential melt is “refrozen” by the decrease in the heat deficit. Storage within the snowpack is tracked in two states: ice (solid) and free water (liquid), the sum of which is termed Snow Water Equivalent ( $SWE$ ), and quantifies the volume of water obtained if all snow is melted. Melt decreases the ice storage of the snowpack, and once the amount of free water surpasses the maximum capacity, water exits the snowpack. Sublimation loss from the snowpack occurs according to the following equation:

$$B_h = \xi \cdot ET_h \cdot f sca_h \quad (2.26)$$

where  $ET_h$  is the potential evapotranspiration described in 2.1.3 The process lowers the non-isothermal heat deficit by:  $Tpk_h \cdot ET_h \cdot 1.27$  calories.

The evolution of the snow depth ( $D$ ) is approximated from the ordinary differential equation [1]:

$$\frac{dD_h}{dt} + \tau \cdot D_h = \frac{Pns_h}{\rho_{init}} + \left( \tau \cdot \frac{SWE_h + Pns_h}{\rho_{max}} \right) \quad (2.27)$$

and is calculated as:

$$\Delta D_h = \Delta t \cdot \left( \frac{Pns_h}{\rho_{init}} + \tau \cdot \left( \frac{SWE_h + Pns_h}{\rho_{max}} - D_h \right) \right) \quad (2.28)$$

Where  $\rho_{init}$  and  $\rho_{max}$  are parameters for initial and maximum density of new snowfall, respectively.  $Pns_h$  is the net new snowfall after interception. The snowpack density is then:

$$\rho_h = \frac{SWE_h}{D_h} \quad (2.29)$$

The snow-covered area is determined by a multi-modal depletion curve [1] that describes the evolution of the fractional snow-covered area ( $f_{sca}$ ) in function of the fraction of maximum  $SWE$ . Parameter  $A_{curve}$  specifies the 11 values of the curve. The depletion curve is used when HRU  $SWE$  is less than the maximum  $SWE$  corresponding to total snow cover specified by parameter  $SWE_{max}$ .

Water available at the soil surface (from melt and rain) proceed to both infiltrating the soil reservoir and filling the impervious zone reservoir.

### 2.1.5 Impervious Zone Reservoir

The impervious zone reservoir ( $Simp$ ) constitutes the fraction of the HRU specified by the parameter  $f_i$  with no soil-infiltration capacity. It receives  $f_i$  fraction of the total water from snow melt ( $M$ ) and rain throughfall ( $Pnr$ ) for each HRU and has maximum retention capacity of  $Simax$ , above which water flows directly as surface runoff  $Fsri$  described in 2.1.7. The process is described in the following discretized dynamic equation:

$$Simp = Simp_{t-1} + (Pnr_h + M_h) \cdot f_{i_h} - Fsri_h - ei_h \quad (2.30)$$

where  $ei$  is the water depleted by evaporation:

$$ei_h = ETa_h \cdot (1 - f_{sca_h}) \cdot f_{i_h} \quad (2.31)$$

where  $ETa_h$  is the available potential for evapotranspiration remaining in the system after accounting for evapotranspiration and sublimation losses that already occurred (interception,...).

### 2.1.6 Soil Zone Reservoir

The soil zone reservoir ( $Ssz$ ) represents the active soil profile and its depth is considered to be the average rooting depth of the predominant vegetation on the HRU. It receives part of the water from snow melt and rain. It can hold up to the parameter  $Sszmax$ . The upper layer is termed the recharge zone ( $Ssre$ ) and the lower layer is termed the lower zone ( $Ssl$ ). While water is lost through both evaporation and transpiration from the upper layer, it is only lost through transpiration from the lower layer. Evaporation is limited by both how much potential of  $ET$  is left in  $ETa_h$  for the day and by the availability of water. Water needs to fill the recharge layer before it can proceed to the lower layer. The soil water content represented by  $Ssz$  is also defined as the “soil moisture”. A maximum of  $Fzgwmax_h$  percolates from  $Ssz$  to  $Sgw$ , and the remainder to  $Sss$  as excess.

### 2.1.7 Surface Runoff

Both the pervious soil zone and the impervious reservoirs contribute to surface runoff  $Fsr$  as  $Fsrp$  and  $Fsri$  respectively.

$$Fsr_h = (1 - fi_h) \cdot Fsrp_h + fi_h \cdot Fsri_h \quad (2.32)$$

The pervious runoff is modeled using the following non-linear parametric equation in function of soil moisture, net rain precipitation and snowmelt [1]:

$$Fsrp_h = max \left[ \alpha \mathbf{1}_h \cdot 10^{\theta \mathbf{1}_h \cdot Ssz \cdot (0.5 \cdot Pn_h)}, \mathbf{A}sr \right] \cdot (Pnr_h + M_h) \quad (2.33)$$

Where  $Pn_h$  is net precipitation (2.14),  $\alpha \mathbf{1}_h$ ,  $\theta \mathbf{1}_h$  and  $\mathbf{A}sr$  are model parameters. The impervious runoff is equal to the amount of water exceeding  $\mathbf{S}imax$  in (2.30).

### 2.1.8 Subsurface Reservoir

The subsurface reservoir models the relatively rapid water movement of the unsaturated soil zones towards a streamflow channel. Such behavior is highlighted typically during and after a rainfall or snowmelt event. A reservoir routing system is used to model the flow from the subsurface reservoir. The flow  $F_{ss}$  is the solution that satisfies both the mass continuity equation:

$$F_{ss_s,t} = F_{ss_s,t-1} - \frac{d(S_{ss})}{dt} \quad (2.34)$$

and the empirical quadratic relation:

$$F_{ss_s,t} = \alpha_{3_s} \cdot S_{ss_s} + \beta_{3_s} \cdot S_{ss_s}^2 \quad (2.35)$$

where  $\alpha_{2_s}$  and  $\beta_{2_s}$  are routing coefficients.

Another  $S_{ss}$  discharge route transfers water to the  $S_{gw}$  reservoir through the parametric power equation:

$$F_{ss\_gw_s} = \alpha_{2_s} \cdot \left[ \frac{S_{ss_s}}{smax_s} \right]^{\theta_{2_s}} \quad (2.36)$$

where  $\alpha_{2_s}$ ,  $smax$  and  $\theta_{2_s}$  are model parameters.

### 2.1.9 Ground-Water Reservoir

Once the soil zone reservoir is saturated, excess water starts filling the ground-water ( $S_{gw}$ ) reservoir at a maximum daily recharge rate  $Fzgwmax_h$ , a parameter defined for each HRU.

$$F_{sz\_gw} = \min(Fzgwmax_h, S_{sz\_excess}) \quad (2.37)$$

Water leaving the ground-water reservoir either flows laterally ( $F_{gw}$ ) contributing to the baseflow portion of the streamflow or is lost from the system via the groundwater sink, both at a linear rate:

$$F_{gw_h} = \alpha_{4_g} \cdot S_{gw_g} \quad (2.38)$$

$$F_{gsnk_h} = \alpha_{5_g} \cdot S_{gw_g} \quad (2.39)$$

### 2.1.10 Model Output: Streamflow

Basin streamflow, which constitutes the model output, is computed as an area weighted average of all the reservoir flow outputs described above:

$$F_{basin} = \sum_{h=1}^{N_h} \frac{Fsr_h}{A_h} + \sum_{s=1}^{N_s} \frac{Fss_s}{A_s} + \sum_{g=1}^{N_g} \frac{Fgw_g}{A_g} \quad (2.40)$$

Improving the daily forecast accuracy of  $F_{basin}$  is the ultimate objective of this report.

## 2.2 Model Features

As can be seen by equations in section 2, the model is extremely nonlinear (ex: equations 2.22, 2.33, and 2.36...) and there is frequent use of thresholds to determine what situational equations are appropriate to use (ex: equation 2.13). Those threshold functions also make the resulting system not continuously differentiable.

There are some explicit constraints for model inputs, states and parameters shown in tables 2.1, 2.2 and A.1. These constraints must be satisfied for the model to run and to get a physically realistic behavior.

The model is spatially distributed and runs independently on each sub-basin. Each sub-basin is divided into HRUs (for surface), SSRs and GWRs (for sub-surface) spatial units. The basin-level model dimension depends on the number of these sub-units.

To quantify the model dimension in this study, we focus on the region of interest, which is the East Branch of the North Fork of the Feather River sub-basin, highlighted in Fig. 2.1. 111 HRUs have been apriori selected and they are spatially co-located with the ground-layer subsurface reservoirs (SSRs) and groundwater reservoirs (GWRs). We will thus call the 111 regions HRUs in the remainder of this study. This implies that  $A_h = A_s = A_g$  in (2.40). Each HRU has a total of 12 states, shown in Table 2.2. In our case, model parameters can be basin-wide (i.e. fixed and constant value for all HRUs) or different for each HRU. Some parameters are constant and others change monthly or seasonally.

The dimension of the model in the region of interest is then 1332.

Data In this study, model inputs are generated by various cooperators and are obtained from Pacific Gas & Electric (PG&E). They were estimated using the Parameter-Elevation Relationships on Independent Slopes Model (PRISM)'s historical monthly maps [8] and daily precipitation gauges. More details are available in [7].

SWE measurements are simulated by averaging for each HRU from the 90 meter resolution daily historical product in [9].

Other measurements available are the streamflow  $F_{basin}$  in equation 2.40 downstream of the basin to check the output performance. The model parameters used are those obtained from the previous calibration study of the model and can be found in [7].

## Chapter 3

# Filtering Techniques

### 3.1 Overview

Models are simplifications of physical processes and are inherently imperfect representations of reality. Combining related measurements with modeled estimates through data assimilation would produce more accurate outcomes. Assimilation approaches were introduced in oceanography and meteorology [10], [11] with different schemes used such as the variational methods [12] and the filtering techniques including nudging [13], optimal interpolation [14] and Kalman filtering with its variants that are presented below.

#### Bayesian Filtering

Bayesian inference [15] is the underlining principle for all the data assimilation approaches we will describe in this section. They stem from the two following probability rules: marginalization (sum) rule: Given  $f_{xy}(x, y)$ , we define:

$$f_x(x) := \sum_{y \in Y} f_{xy}(x, y) \quad (3.1)$$

and the conditioning, or product rule defined as:

$$f_{x|y}(x, y) := \frac{f_{xy}(x, y)}{f_y(y)} \quad (3.2)$$

for  $f_y(y) \neq 0$

We can formulate the following forecast and update system in the Bayesian framework, where  $x$  represents the model state,  $z$  represent the measurements and  $k$  is the discrete time step:

Prior update (or forecast):

$$f(x(k)|z(1:k-1)) = \int f(x(k)|x(k-1)) \cdot f(x(k-1)|z(1:k-1)) dx(k-1) \quad (3.3)$$

where the first item in the integral is the process model and the second item is the previous time step measurement update result.

Measurement update (or update):

$$f(x(k)|z(1:k)) = \frac{f(z(k)|x(k)) \cdot f(x(k)|z(1:k-1))}{\int f(z(k)|x(k)) f(x(k)|z(1:k-1)) dx(k)} \quad (3.4)$$

where the first and second items of the numerator are the measurement model and the prior (3.3), respectively, and the denominator is the normalization  $f(z(k)|z(1:k-1))$ .

Kalman Filter It can be shown [16]–[18] that given the system of equations with a linear process model  $A$  and independent Gaussian distributions for errors  $v(k-1)$  and  $w(k)$  and prior: assuming initialization:  $\hat{x}_a(0) = x_0, P_a(0) = P_0$

$$x(k) := A(k) \cdot x(k-1) + u(k-1) + v(k-1) \quad (3.5)$$

$$y(k) := H(k)x(k) + w(k) \quad (3.6)$$

the analytical solution to the Bayesian state estimation problem in (3.3) and (3.4) is the Kalman Filter (KF) and can be written in a recursive form while keeping track of the distributions' mean and covariance as:

Prior update/Forecast step:

$$\hat{x}_f(k) = A(k-1)\hat{x}_a(k-1) + u(k-1) \quad (3.7)$$

$$P_f(k) = A(k-1)P_a(k-1)A^T(k-1) + Q(k-1) \quad (3.8)$$

A Posteriori update/Analysis step:

$$\hat{x}_a(k) = \hat{x}_f(k) + P_a(k)H^T(k)R^{-1}(k)(z(k) - H(k)\hat{x}_f(k)) \quad (3.9)$$

$$P_a(k) = \left( P_f^{-1}(k) + H^T(k)R^{-1}(k)H(k) \right)^{-1} \quad (3.10)$$

without the need to store all previous observations and states. It is commonly referred to  $P_a(k)H^T(k)R^{-1}(k)$  in (3.9) as the Kalman gain  $K(k)$ . The maximum a posteriori estimate (MAP) of the normally distributed analysis coincides with the mean:  $x_a^{MAP}(k) = \hat{x}_a(k)$  and thus is chosen as the best estimate of the system state at time  $k$ .

The basic Kalman Filter (KF) algorithm that is restricted for only linear models consists of two steps, a predict step followed by an update step. The KF and its variants have numerous applications in the physical sciences, primarily in the guidance [19], navigation and control of vehicles [20], time-series analysis [21] and robotics [22].

#### Extended Kalman Filter

The Extended Kalman Filter (EKF) is an extension to the KF for non-linear systems. The approach is to linearize the non-linear process model around the mean using first order Taylor expansion. It thus becomes an approximation of the exact solution. The Jacobian is then used to advance the error covariance matrix in time. It requires the model to be continuously differentiable and is shown to have bad performance when the system exhibits strong non-linearities. For instance, the EKF was successfully used to integrate GPS-derived flow velocity measurements from drifters to improve the non-linear hydrodynamics state estimates in a controlled channel pilot experiment [23].

#### Unscented Kalman Filter

The Unscented Kalman Filter (UKF), also referred to as Sigma-Point KF, tries to solve the linearity problem using another approach. It works by first deterministically sampling using an algorithm sigma points from the distribution and assigning to each point a weight. At every time step the points are individually propagated through the non-linear model,

and finally the resulting mean and covariance is computed from the forecasted sigma points. Unlike the EKF, not only the mean is advanced in time, but also many points sampled from the distribution. The method requires  $N = 2r + 1$  sigma points where  $r$  is the system dimension. [24] shows that the UKF, with the same complexity, is a better filter than the EKF for non-linear models capturing sigma points' mean and covariance accurately up to the third order (in Taylor series expansion terms) while the EKF only to the first. Using UKF instead of EKF halved the RMSE in estimating vehicle and wheel angular velocity of the anti-lock braking system (ABS) in [25].

#### Ensemble Kalman Filter

The EnKF represents the distribution as a random sample of points called ensemble members. In the forecast step, each ensemble member is integrated forward in time with an additional process noise. The forecasted error covariance matrix is approximated at any time step by the ensemble sample covariance. The error covariance matrix is thus implicitly propagated saving the high computational requirements associated with its storage and forward integration such as in KF and EKF. The update stage consists of updating every ensemble member using sample covariance as an approximation of the true covariance. The ensemble members are never resampled from the distribution thus preserving the ensemble skewness, kurtosis, clustering, etc. In the EnKF update, the updated ensemble is obtained by shifting and re-scaling the forecast ensemble.

#### Particle Filter

The Sequential Importance Resampling (SIR) Particle Filter (PF), samples points called particles from the distribution to generate system states realizations. These particles are assigned weights that represent their likelihood. The sampling is done randomly like the EnKF, but not necessarily from normal distributions. On update, the particle weights are adjusted. Higher weights are assigned to those that are supported by sensor data and vice versa. The Particle Filter suffers from a problem with sample degeneracy and impoverishment, in that some scenarios end up with very few high-weight particles while the other majority of particles have almost zero weights. Another disadvantage is that

the number of particles required for good performance scales exponentially with the model dimension. The method does not require Gaussian distributions nor model linearity [26]. It thus works well with distributions that have complex shape and multiple peaks. Example applications include vehicle localization [27], indoor occupant positioning using a Radio Frequency Identity RFID and receivers [28], etc.

### Filter Selection

The Kalman Filter in its basic form cannot be applied to our modeling system because the latter is non-linear. The model is also not continuously differentiable, which eliminates the option of using the Extended Kalman Filter. For example, different physical conditions can lead to the use of totally different process equations. For our application, using either of UKF, EnKF or PF is theoretically valid. According to [29], UKF would require 2337 sigma points to fully represent the mean and variance of the system. In [30], the UKF was used in a similar application as ours but using algorithms to reduce the number of sigma points to  $n + 2$ . Multiple studies in the literature pertaining to geophysics, oceanography and hydrology have shown that the EnKF performs well even when the ensemble size is orders of magnitude smaller than the system dimension [31], [32]. EnKF outperformed PF in the assimilation of a conceptual rainfall-runoff in [33] and the justification was that it is “less sensitive to misspecification of the model and uncertainties”, and in [34], it had similar performance. Given the above specified reasons and the limitation in resources available that will restrict the number of model realizations, we choose the EnKF.

## **3.2 EnKF & State-space**

The Ensemble Kalman Filter [35] was introduced in [36] as an alternative to EKF to overcome specific difficulties with nonlinear state evolution models, including non-differentiability of the model and closure problems. As described above, EnKF uses Monte Carlo (or ensemble integrations) and has the same standard update equations of the KF, with state mean and covariance replaced by an estimated ensemble sample mean and covariance respectively [37]. The EnKF and more generally statistical inversion schemes were

greatly advanced in [38], mainly in the context of Tomography. A framework for sampling a smooth prior is presented, especially when structural information about the ensemble (correlations between pixels) is known apriori. Such techniques can also be used in the perturbations of spatial data such as climatologic inputs, model error etc. For instance these techniques were used to generate smooth initial spatial model states ensemble in [39] without shocks. [40] shows that using a non-trivial state covariance matrix in the state noise yields much better performance than a stationary one.

With the Kalman Smoother, state estimation was applied in Electrical Impedance Tomography (EIT) to reconstruct rapidly time-varying cross-sectional body scans inferred from boundary voltage and current measurements [41], [42]. Similarly, Electrical Resistance Tomography (ERT) measurements were used to infer water saturation levels in soil using EKF assimilation of a weakly non-linear model [43]. Time-dependent noise was used in [44] infer gas temperature from resistance measurements.

[39] provided a framework for assimilating readily available GPS measurements of vehicle speed and velocity into a traffic model [45] using the EnKF as part of the the Mobile Millennium traffic-monitoring system [46]. Results of an unprecedented scale experiment [39], where GPS speed and position data from 100-vehicles were assimilated into the traffic model, showed improved estimates of the traffic state even with very low percentage of GPS-equipped vehicles participating.

We now present the conceptual algorithm of the EnKF: Let  $s$  be the system dimension. Hydrologic models typically take inputs. Let  $i$  be the number of inputs and  $o$  be the number of observations. Let  $\mathbf{x}$  be the state vector and  $\mathcal{M}$  the non-linear model. The discrete time  $k$  forward integration equation can be represented as:

$$\mathbf{x}^k = \mathcal{M}(\mathbf{x}^{k-1}, \mathbf{b}^k) \tag{3.11}$$

Where the model observed states are:

$$\mathbf{y}^k = \mathbf{H}\mathbf{x}^k \tag{3.12}$$

The localized EnKF assimilation scheme can be summarized by the algorithm below:

1. *Initialization*: Draw  $N$  ensemble realizations  $\mathbf{x}_a^0(e)$  (with  $e \in [1, \dots, N]$ ) from a process with a mean  $\bar{\mathbf{x}}_a^0$  and covariance  $\mathbf{C}_a^0$
2. *Forecast*: At each time step  $n$ , update each of the  $N$  ensemble members according to PRMS forward simulation model. Then update the ensemble mean and covariance according to:

$$\mathbf{x}_f^k(e) = \mathcal{M}[\mathbf{x}_f^{k-1}(e), \mathbf{b}^k(e)] + \boldsymbol{\eta}^k(e) \quad (3.13)$$

$$\bar{\mathbf{x}}_f^k = \frac{1}{N} \sum_{e=1}^N \mathbf{x}_f^k(e)$$

$$\mathbf{P}_{ens,f}^k = \frac{1}{N-1} \sum_{e=1}^N \left( \mathbf{x}_f^k(e) - \bar{\mathbf{x}}_f^k \right) \left( \mathbf{x}_f^k(e) - \bar{\mathbf{x}}_f^k \right)^T$$

3. *Analysis*: Assuming a measurement is obtained at time  $n$ , localize sample covariance, compute the Kalman gain, and update the network forecast:

$$\mathbf{P}_{ens,f,L}^k = \mathcal{L} \circ \mathbf{P}_{ens,f}^k$$

$$\mathbf{K}_{ens}^k = \mathbf{C}_{ens,f,L}^k \left( \mathbf{H}^k \right)^T \left( \mathbf{H}^k \mathbf{C}_{ens,f,L}^k \left( \mathbf{H}^k \right)^T + \mathbf{C}_{obs}^k \right)^{-1}$$

$$\mathbf{x}_a^k(e) = \mathbf{x}_f^k(e) + \mathbf{K}_{ens}^k \left( \mathbf{y}_{obs}^k - \mathbf{H}^k \mathbf{x}_f^k(e) + \boldsymbol{\chi}^k(e) \right) \quad (3.14)$$

4. Return to 2.

where:

$\mathbf{s}_f^k$  is the ensemble of states forecast at time  $k$ , [s x N]

$\mathcal{M}$  is the model operator (PRMS)

$\mathbf{b}^k$  is the perturbed ensemble input, [i x N]

$\boldsymbol{\eta}^k$  is the white process noise with covariance  $\mathbf{Q}$ , [s x N]

$\bar{\mathbf{s}}_f^k$  is the ensemble sample mean of the states forecast at time  $k$ , [s]

$\mathbf{P}_{ens,f}^k$  is the ensemble states forecast sample covariance at time  $k$ , [s x s]

$\mathbf{H}^k$  is the observation matrix at time k, [o x s]

$\mathbf{K}_{ens}^k$  is the Kalman gain, [s x o]

$\mathbf{R}_{obs}^k$  is the measurement error covariance, [o x o]

$x_a^k$  is the ensemble states after analysis at time k, [s x N]

$\mathbf{y}_{obs}^k$  is the vector of observations, [o x 1]

$\chi^k$  is the measurement perturbations with 0 mean & covariance  $\mathbf{R}_{obs}^n$ , [o, N]

One can refer to [47] for more details and for practical implementation.

### Model Constraints

The model described above has physical constraints for inputs, states, observations and parameters as described in section 2.2. During the perturbation of such quantities with normally distributed noise and after the update step of the EnKF, invalid scenarios are likely to arise. To circumvent this issue, a hard boundary check is implemented after each analysis step. Any values outside the range are set to the corresponding maximum or minimum of that state.

Major determinants of EnKF performance are the ensemble size and the error statistics.

### Ensemble Size

With an infinite ensemble size, the EnKF approximation will reach the true KF solution. We choose an ensemble size of 100 for practical reasons detailed in Section 4.

Additionally due to the small ensemble size compared to the state space size, it might be necessary to use localization when using the EnKF when spurious correlations between physically uncorrelated states (ex: geographically distant states) affect the filter performance [32], [48]. Distance based localization weights using linear, exponential or Gaussian decorrelation functions are typically used to attenuate the correlations between distant states and observations. For this study, we do not use any localization as the impact on the outflow is assumed to be negligible compared to the scheme introduced in section 4.2.

### Inputs and Model Error Statistics

The additive noise  $\boldsymbol{\eta}^k$  in the integration equation at time  $k$  (3.13) is modeled as a fraction of the mean of the ensemble. This scheme was chosen because the majority of states are storage states and have zero lower bound. Simultaneously forcing such boundaries as described above and adding daily zero mean normal noise would introduce bias to the model. The downside of such error modeling is that the model is assumed to have perfect estimates for states with zero means.

$$\boldsymbol{\eta}_k \sim \mathcal{N}(\mu = 0, \sigma = \alpha \bar{\mathbf{x}}_f^k) \quad (3.15)$$

Inputs (and potentially parameters) perturbation propagates model errors that are consistent with the physics of the model, compared to only a simple daily addition of white noise to the ensemble. Inputs and observations error perturbations are modeled as independent normal distributions with standard deviations computed as the following:

For input precipitation:

$$\sigma_P = 0.4 \cdot P \quad (3.16)$$

$$Pens \sim \mathcal{N}(\mu = P, \sigma^2 = (0.4 \cdot P)^2) \quad (3.17)$$

For simulated SWE measurement:

$$\sigma_{SWE} = 0.1 \cdot SWE \quad (3.18)$$

The standard deviation for temperature is assumed to be a constant:

$$\sigma_T = 2^\circ C \quad (3.19)$$

$$Tens \sim T + \mathcal{N}(\mu = 0, \sigma^2 = 2^2) \quad (3.20)$$

Variance inflation is required when the model forecast error is known to be underestimated. Without variance inflation, the filter tends to diverge and the observation fails to influence the model forecast because of the underestimated spread in the ensemble state

forecast due to previous update events. Multiple studies implement variance inflation differently [48] and some use adaptive inflation that is pre-computed at analysis step as a function of the error between the forecasted state and the observation [49]. In this study we will use a post-analysis inflation procedure with constant  $\alpha_i = 0.9$  similar to [50].

$$\mathbf{x}'_{a,infl} = (1 - \alpha_i) \mathbf{x}'_a + \alpha_i \cdot \mathbf{x}'_f \quad (3.21)$$

where the prime  $\mathbf{x}'$  denotes the ensemble anomaly ( $\mathbf{x} - \bar{\mathbf{x}}$ ).

## Chapter 4

# Experimental Results

The average water year 2006 was chosen for simulation. Additional water years (2011, 2014) are simulated and their results can be found in Appendix B. 2014 was a dry year while 2011 was a wet water year [51]. We first run the model with the calibrated parameters [7] which are presently used operationally without any assimilation to serve as a reference. Next, we proceed to assimilate the SWE state of the model in Section 4.1. The experiment results motivate a new joint-state parameter assimilation scheme with feedback that is explained and simulated in Section 4.2.

### Performance Metrics

The Root Mean Squared Error is used to compare the stream flow model output from the assimilation experiments with the measured stream flow.

$$RMSE = \frac{1}{T} \sum_{t=1}^T \sqrt{(Fbasin_{meas,t} - Fbasin_{sim,t})^2} \quad (4.1)$$

The RMSE of all experiments conducted are summarized in tables 4.2 and 4.3.

For ensemble size, we use the maximum possible ensemble number of 100 that is practical to run on a typical user accessible computer. The limiting factor for increasing the ensemble size in our case is the time needed to read and write state values to an ensemble of PRMS files. The PRMS used is available as a pre-compiled binaries of a Fortran code and is setup to read/write state data from/to files. To improve the latency of these procedures, the

files were stored on RAM instead of disk and the ensemble model forecasts were generated in parallel in a multi-threaded wrapper. Simulation time for one water year with daily assimilation is around one hour.

## 4.1 SWE State Assimilation

Figure 4.2 shows that the analysis SWE follows closely the “measured” SWE. This is expected since inflation was used to increase the uncertainty of the SWE state forecast, and thus daily observations of SWE have a relatively smaller uncertainty during the majority of the simulation days. The remaining water year graphs are available in Appendix B. Since the SWE observation uncertainty was modeled as a percentage of the observation value, it peaks during the peak snow season and is illustrated by the grey ensemble spread in the plot.

Figure 4.1 shows a deterioration in the performance in terms of runoff output. This is consistent for all simulation years, as indicated in Table 4.2 and in appendix B. The main explanation for such outcome is that the model parameters used are those that were calibrated based on inputs and output only. The former are temperature and precipitation while the latter is the measured runoff. No SWE data were involved in the calibration, which implies that updating the model SWE to a more accurate estimate will not necessarily improve the model output. We should also note that the SWE used as observation is not directly measured, but a modeled product. However, we assume this is not the reason of deterioration. Furthermore, the runoff output ensemble spread is not as large as desired, since the measured runoff is not within the ensemble for the majority of the days.

Given that the SWE ensemble spread is relatively large, that the parameters used were calibrated without SWE knowledge, we conclude that it would be advantageous to perturb the model parameters and estimate them as well in a joint state-parameters assimilation scheme. Moreover a feedback consisting of the previous-day measured runoff would be necessary to estimate those parameters mostly involved in runoff generation dynamics.

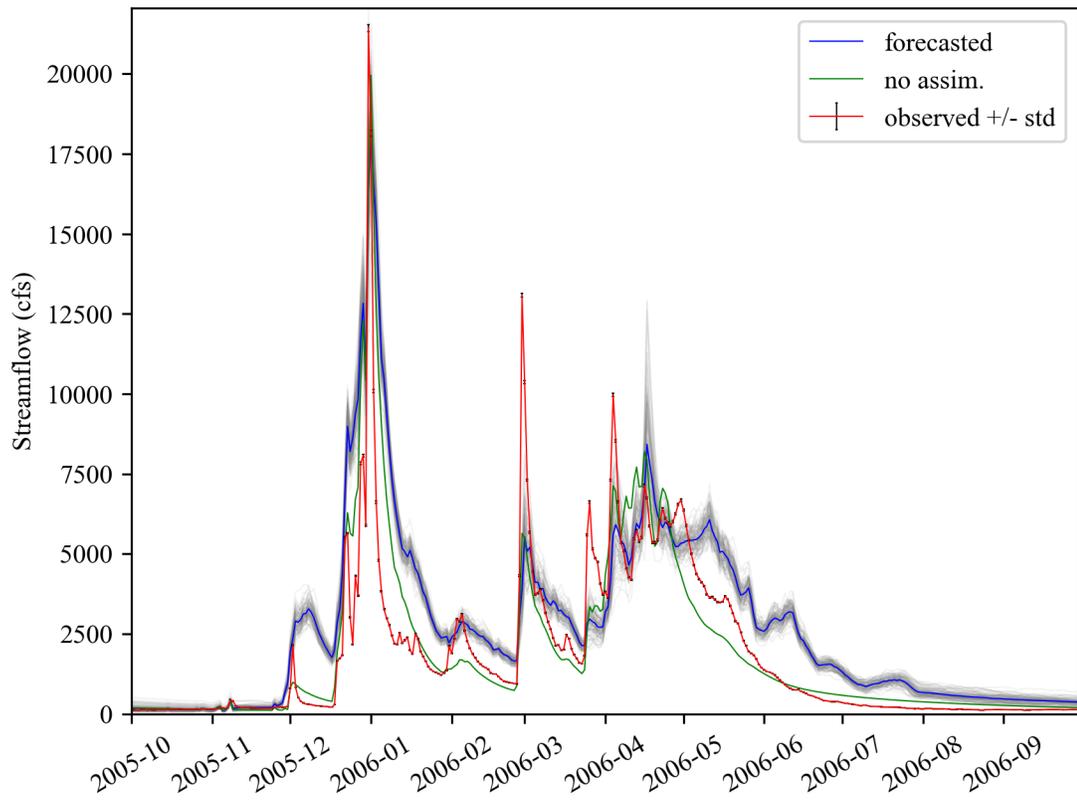


Figure 4.1: Water year 2006 basin streamflow output.

## 4.2 SWE and Runoff Joint State-Parameter Assimilation

### 4.2.1 New State Space

Results of the previous experiment imply that parameters need to be altered. Model parameters can be appended to the state vector and updated online on assimilation events in what is termed a “Joint state-parameter estimation”. In such method, there is no need to apriori train the model to find the optimal parameters which might not be time-independent. Instead, the range of each parameter must be known. Parameters are indirectly updated through the cross-correlation between the parameter and the “state” being observed in the Kalman gain matrix.

Joint parameter-state assimilation has been used in many applications with complex models and spans different fields to name a few: assimilation of measured blood concentration into a metabolic model to update the numerous model parameters as well as other unobserved states [52], assimilation of measured velocity fields from GPS-equipped drifters into a shallow water model to update model parameters [53], and assimilation of measured displacements into gas storage geomechanical model to reduce uncertainty of some model parameters [54]. Results in [43] showed that simultaneous parameter estimation is necessary when the model perturbed by daily additive white noise is far from the truth.

Parameters involved in model dynamics from model states to streamflow output require streamflow measurements to be updated. Parameters that are included in the state space were chosen based on the sensitivity results obtained from the study in [55]. We choose the parameters that score high on daily streamflow statistics of mean, CV and AR1 indicating sensitivity of parameters to each process. Processes chosen were snowmelt (assimilating SWE) and runoff (assimilating runoff), with emphasis on runoff. Chosen parameters include 9 parameters per HRU and 4 global parameters and are shown in Table 4.1. The state vector is augmented with these parameters. Exponential parameters are excluded (ex: `smidx_exp`) to maintain approximate Gaussian ensemble distributions, a requirement of the EnKF.

Symbol	PRMS symbol	Quantity	Type	Range	Units	Description
Tms	tmax_allsnow	1	constant	-10, 40	°F	temperature below which P is all snow
Tmr	tmax_allrain	1	monthly	-8, 60	°F	temperature above which P is all rain
$\alpha 1$	smidx_coef	per HRU	constant	0.001, 0.06	-	linear surface runoff (Fsr) coefficient
Asr	careas_max	per HRU	constant	0.0, 1.0	-	max areal fraction contributing to surface runoff Fsr
$\beta$	dday_intcp	1	monthly	-60.0, 10.0	°day	intercept of degree-day equation
jc	jh_coef	1	monthly	0.005, 0.06	/°F	coef. used in ETp
Sszmax	soilmoist_max	per HRU	constant	0.001, 60.0	inches	maximum soilzone water holding capacity
$\alpha 4$	gflow_coef	per GWR	constant	0.001, 0.5	-/day	linear coefficient routing Sgw to streamflow
Fzgwmax	soil2gw_max	per HRU	constant	0.0, 5.0	inches	max soil excess water routed to gwStr
$\alpha 5$	gwsink_coef	per GWR	constant	0.0, 1.0	-/day	linear coefficient for groundwater sink
$\alpha 2$	ssr2gw_rate	per SSR	constant	0.05, 0.8	-/day	linear coefficient routing Sss to Sgw
$\alpha 3$	ssrcoef_sq	per SSR	constant	0.0, 1.0	-	coefficient routing Sss to streamflow
$\beta 3$	ssrcoef_lin	per SSR	constant	0.0, 1.0	-/day	linear coefficient routing Sss to streamflow

Table 4.1: Perturbed and updated parameters.

Initial parameter perturbations are modeled proportionally to their range, such that:

$$\sigma_p = 0.25 \cdot |p_{max} - p_{min}| \quad (4.2)$$

Parameters are inflated post-analysis similarly to states as described in (3.21). Parameters do not change during model integration, thus when the parameter ensemble nears collapse - ie. when the standard deviation becomes less than  $\sigma_{target}$ , inflation is performed by re-scaling the parameter ensemble perturbations to  $\sigma_{target}$ :

$$\sigma_{target} = 0.25 \cdot \sigma_p \quad (4.3)$$

where  $\sigma_p$  is the standard deviation of the initial parameter perturbation.

Next, we additionally augment the state-space with the previous-day updated runoff, so that it can be updated on the next analysis day by the previous-day measured runoff. The current day simulated streamflow is also appended. Finally, the dimension of the augmented state-space becomes 2337.

The observation vector  $y_{obs}$  is augmented by the previous-day measured runoff and correspondingly, a new row is appended to the observation matrix  $H$  in the analysis equation (3.13). The previous-day measured runoff is available in practical applications, because we are updating the SWE state daily. The measured streamflow is assumed near perfect and perturbed with an independent normally distributed error with a scaling standard deviation:

$$\sigma_F = 0.005 \cdot F_{basin_{meas}} \quad (4.4)$$

Water Year	no assim.	SWE update	$\Delta(\%)$	Joint SWE/Fbasin	$\Delta(\%)$	AR(1)	$\Delta(\%)$
2006	1228	1845	50	889	-28	1235	1
2011	870	1891	117	360	-59	430	-51
2014	198	996	404	177	-10	126	-36

Table 4.2: Experiments performance in terms of streamflow output RMSE

Water Year	no assim.	SWE update	$\Delta(\%)$	Joint SWE & Stream	$\Delta(\%)$
2006	1.08	0.02	-98	0.04	-96
2011	1.97	0.05	-97	0.05	-98
2014	0.20	0.01	-93	0.01	-95

Table 4.3: Experiments validation in terms of basin-mean SWE RMSE.

### 4.2.2 Results

As reference, we compare the results to both the estimated streamflow with no assimilation, which is the green plot shown in Figure 4.3, and an additional estimate which is simply the previous-day measured streamflow (not visualized), also termed autoregressive AR(1) filter with equation:

$$F_{basin_{sim},t} = F_{basin_{meas},t-1} \quad (4.5)$$

Results of both experiments are summarized in Tables 4.2 and 4.3 and show a consistent improvement in simulated streamflow compared to the no assimilation parameter-calibrated case, with up to 60% reduction in RMSE for the wet water year 2011. Results were occasionally better than the naive previous-day AR(1) forecast such as the average water year 2006 (28%), but worse during the dry water year 2014. For all assimilation experiments, the RMSE of the basin-mean SWE state was substantially reduced (> 90%) as expected, since it is the state being “measured” with 10% uncertainty. SWE results constitute a validation of the EnKF analysis procedure.

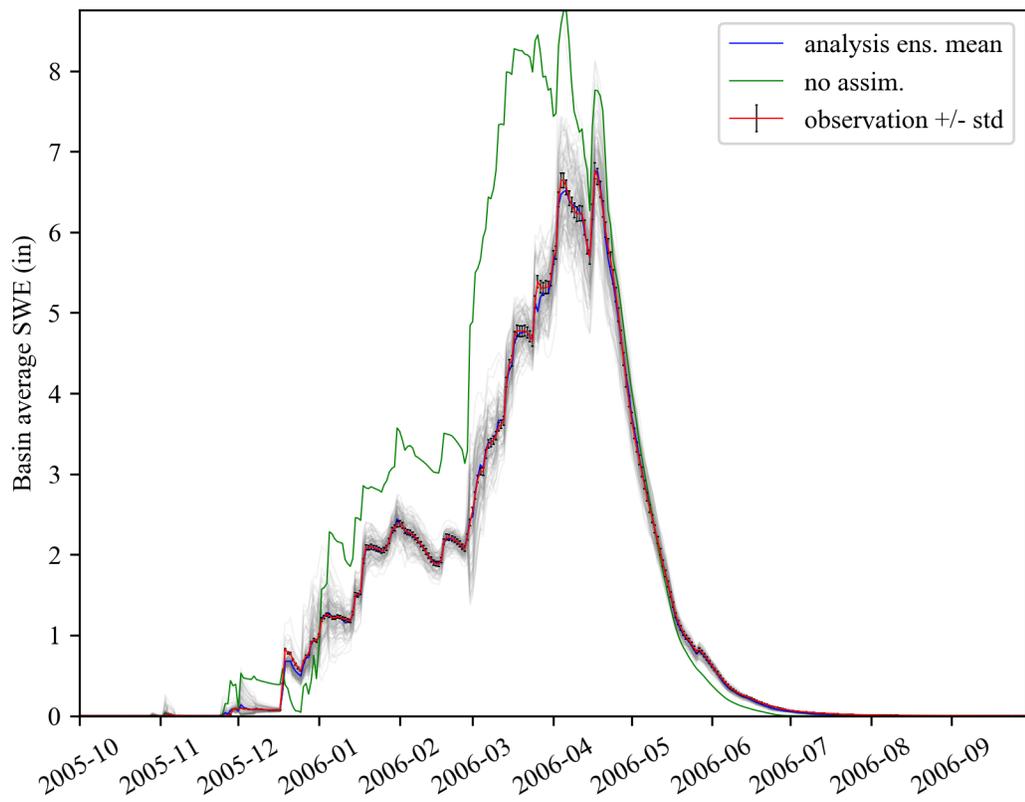


Figure 4.2: Water year 2006 basin average SWE.

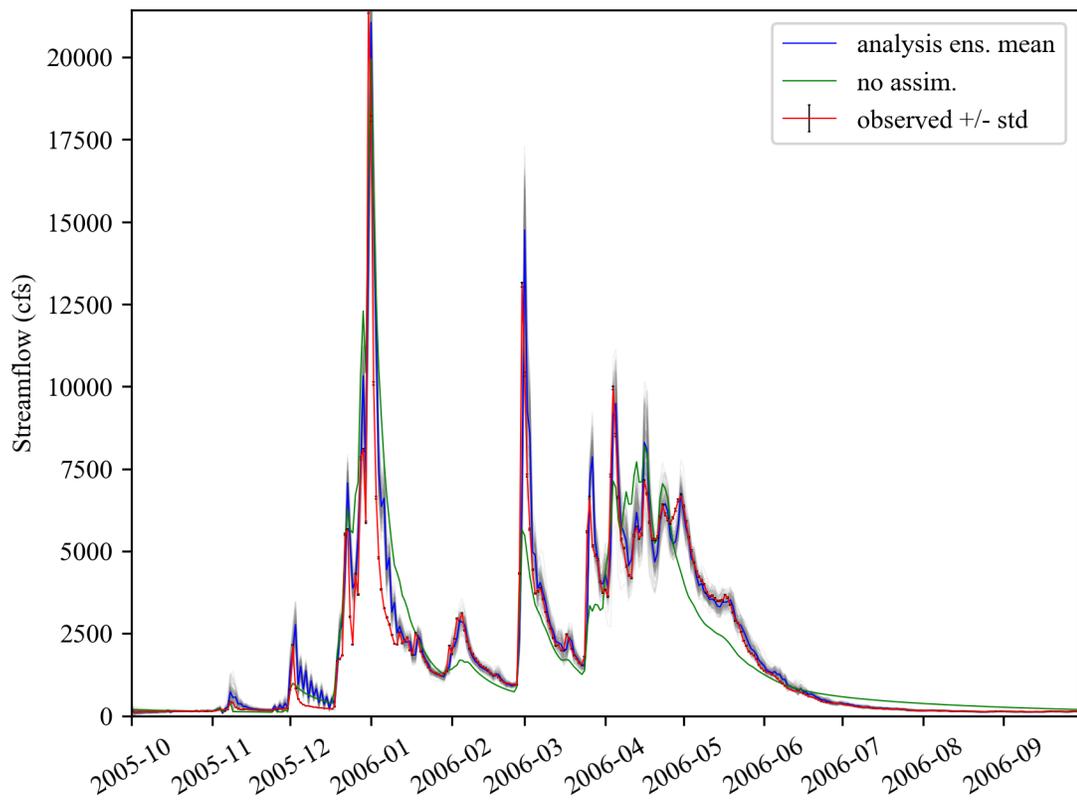


Figure 4.3: Water year 2006 basin average streamflow.

## Chapter 5

# Conclusion

The EnKF data assimilation framework presented succeeded in updating the modeled SWE during analysis step. SWE state assimilation alone does not improve runoff in a heavily parameterized model, where the parameters have been calibrated based on inputs/output without taking into consideration the SWE state being updated. Furthermore, results show that deterioration in accuracy can occur. We postulate SWE assimilation in a non-parametric model is likely to improve streamflow forecast accuracy.

Joint state-parameter assimilation using the previous day measured stream flow as feedback shows a substantial improvement in the accuracy of the daily estimate of streamflow (30% reduction in RMSE for water year 2006) compared to the calibrated no-assimilation scenario, and compared to the simple previous-day AR(1) estimator (30% reduction in RMSE for water year 2006) during average water years.

In reality, streamflow measurements are not as accurate as assumed in this study. It is typically computed from the river stage measurements assuming constant cross section area and other approximations that should be accounted for. Streamflow measurement uncertainty, if well known, can be modeled and accounted for in 4.4. Nevertheless, what this study shows is that with a near-perfect measurement of streamflow, the EnKF framework presented improves the accuracy of the streamflow estimate. More accurate measurement

methods do exist such as current meters or the “Acoustic Doppler Current Profiler”, but are more expensive.

Although the previous day estimator (AR(1)) was better for the dry year 2014, it has disadvantages over using the physical model-based assimilation framework in that it is not a reliable method, especially when the previous streamflow measurement is not available or severely inaccurate due to sensor failure or environmental events, whereas the EnKF framework would provide a more robust and physically sound estimate when measurements are missing. In fact, the only change required when previous day measurement of streamflow is not available is to delete one row from the observation matrix. Moreover, the EnKF with PRMS framework is preferred to the naive AR(1) method given the forecast time window the former can potentially achieve. We suggest that a more sophisticated fully data-driven model that considers at least input data is required to match its potential long-term forecast accuracy such as an artificial recurrent neural network (ex: Long short-term memory, LSTMs).

We suggest as future work to replicate the study for all years where data is available. A longer forecast window study is also of interest to stakeholders and would be a natural continuation of this report, where multiple days in the future are first simulated with model inputs from weather predictions (forecast step), after which their streamflow outputs are updated with measurements of SWE and streamflow available at the current day (update step). This would increase the state dimensions only by the number of days in the prediction window, keeping the assimilation tractable. With parameters updated by measurements of streamflow and states such as SWE, current-day measurements of SWE (and potentially other states) should have a strong impact on the streamflow prediction accuracy for a long prediction window.

## References

- [1] S. L. Markstrom, R. S. Regan, L. E. Hay, and R. J. Viger, “Prms-iv, the precipitation-runoff modeling system, version 4.” [Online]. Available: <https://pubs.usgs.gov/tm/6b7/>
- [2] California Department of Water Resources, “Climate Change,” 2017. [Online]. Available: <http://www.water.ca.gov/climatechange/>
- [3] R. Bales, N. P. Molotch, T. H. Painter, M. D. Dettinger, R. Rice, and J. Dozier, “Mountain hydrology of the western United States,” *Water Resources Research*, vol. 42, p. W08432, 2006.
- [4] I. T. Stewart, D. R. Cayan, and M. D. Dettinger, “Changes in snowmelt runoff timing in western north america under a ‘business as usual’ climate change scenario,” *Climatic Change*, vol. 62, no. 1, pp. 217–232, Jan 2004. [Online]. Available: <https://doi.org/10.1023/B:CLIM.0000013702.22656.e8>
- [5] —, “Changes toward earlier streamflow timing across western north america,” *Journal of Climate*, vol. 18, no. 8, pp. 1136–1155, 2005. [Online]. Available: <https://doi.org/10.1175/JCLI3321.1>
- [6] T. H. Painter, D. F. Berisford, J. W. Boardman, K. J. Bormann, J. S. Deems, F. Gehrke, A. Hedrick, M. Joyce, R. Laidlaw, D. Marks, C. Mattmann, B. McGurk, P. Ramirez, M. Richardson, S. M. K. Skiles, F. C. Seidel, and A. Winstral, “The Airborne Snow Observatory: Fusion of scanning lidar, imaging spectrometer, and physically-based modeling for mapping snow water equivalent and snow albedo,” *Remote Sensing of Environment*, vol. 184, pp. 139–152, 2016. [Online]. Available: <http://dx.doi.org/10.1016/j.rse.2016.06.018>
- [7] K. M. Kocot, A. E. Jeton, Bruce, McGurk, and M. D. Dettinger, “Precipitation-runoff processes in the feather river basin, northeastern california, and streamflow predictability, water years 1971-97.” [Online]. Available: <https://pubs.usgs.gov/sir/2004/5202/>
- [8] C. Daly, M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris, “Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous united states,” *International Journal of Climatology*, vol. 28, no. 15, pp. 2031–2064. [Online]. Available: <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.1688>
- [9] S. A. Margulis, G. Corts, M. Giroto, and M. Durand, “A landsat-era sierra nevada snow reanalysis (19852015),” *Journal of Hydrometeorology*, vol. 17, no. 4, pp. 1203–1221, 2016. [Online]. Available: <https://doi.org/10.1175/JHM-D-15-0177.1>
- [10] R. A. Anthes, “Data assimilation and initialization of hurricane prediction models,” *Journal of the Atmospheric Sciences*, vol. 31, no. 3, pp. 702–719, 1974. [Online]. Available: [https://doi.org/10.1175/1520-0469\(1974\)031<0702:DAAIOH>2.0.CO;2](https://doi.org/10.1175/1520-0469(1974)031<0702:DAAIOH>2.0.CO;2)
- [11] F. COUNILLON and L. BERTINO, “Ensemble optimal interpolation: multivariate properties in the gulf of mexico,” *Tellus A*, vol. 61, no. 2, pp. 296–308. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0870.2008.00383.x>
- [12] F.-X. L. DIMET and O. TALAGRAND, “Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects,” *Tellus A*, vol. 38A, no. 2, pp. 97–110. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0870.1986.tb00459.x>
- [13] C. Paniconi, M. Marrocu, M. Putti, and M. Verbunt, “Newtonian nudging for a richards equation-based distributed hydrological model,” *Advances in Water Resources*, vol. 26, no. 2, pp. 161 – 178, 2003. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0309170802000994>
- [14] A. T. Lorenzo, M. Morzfeld, W. F. Holmgren, and A. D. Cronin, “Optimal interpolation of satellite and ground data for irradiance nowcasting at city scales,” *Solar Energy*, vol. 144, pp. 466 – 474, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0038092X17300555>
- [15] M. Bayes and M. Price, “An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, f. r. s. communicated by mr. price, in a letter to john canton, a. m. f. r. s.” *Philosophical Transactions (1683-1775)*, vol. 53, pp. 370–418, 1763. [Online]. Available: <http://www.jstor.org/stable/105741>
- [16] R. Kalman, “A new approach to linear filtering and prediction problems,” *Journal of Basic Engineering (ASME)*, vol. 82D, pp. 35–45, 01 1960.
- [17] N/A, University Lecture, 2018.
- [18] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*. Springer, 2010.
- [19] J. Fonseca, R. Oliveira, J. Abreu, E. Ferreira, and M. Machado, “Kalman filter embedded in fpga to improve tracking performance in ballistic rockets,” 04 2013, pp. 606–610.
- [20] R. Kasper and S. Schmidt, “Sensor-data-fusion for an autonomous vehicle using a kalman-filter,” in *2008 6th International Symposium on Intelligent Systems and Informatics*, Sep. 2008, pp. 1–5.

- [21] T. Choudhry and H. Wu, "Forecasting the weekly time-varying beta of uk firms: Garch models vs. kalman filter method," *The European Journal of Finance*, vol. 15, no. 4, pp. 437–444, 2009. [Online]. Available: <https://doi.org/10.1080/13518470802604499>
- [22] P. Zarchan and H. Musoff, *Fundamentals of Kalman filtering: a practical approach*. American Inst. of Aeronautics and Astronautics, 2005.
- [23] A. Tinka, M. Rafiee, and A. M. Bayen, "Floating sensor networks for river studies," *IEEE Systems Journal*, vol. 7, no. 1, pp. 36–49, March 2013.
- [24] E. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," vol. 153-158, 02 2000, pp. 153 – 158.
- [25] D. Kun, L. Kaijun, and X. Qunsheng, "Application of unscented kalman filter for the state estimation of anti-lock braking system," in *2006 IEEE International Conference on Vehicular Electronics and Safety*, Dec 2006, pp. 130–133.
- [26] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, Feb 2002.
- [27] A. U. Peker, O. Tosun, and T. Acarman, "Particle filter vehicle localization and map-matching using map topology," in *2011 IEEE Intelligent Vehicles Symposium (IV)*, June 2011, pp. 248–253.
- [28] K. Weekly, H. Zou, L. Xie, Q. Jia, and A. M. Bayen, "Indoor occupant positioning system using active rfid deployment and particle filters," in *2014 IEEE International Conference on Distributed Computing in Sensor Systems*, May 2014, pp. 35–42.
- [29] Y. Cheng and Z. Liu, "Optimized selection of sigma points in the unscented kalman filter," in *2011 International Conference on Electrical and Control Engineering*, Sept 2011, pp. 3073–3075.
- [30] P. Jiang, Y. Sun, and W. Bao, "State estimation of conceptual hydrological models using unscented Kalman filter," *Hydrology Research*, vol. 50, no. 2, pp. 479–497, 12 2018. [Online]. Available: <https://doi.org/10.2166/nh.2018.038>
- [31] Zhang, Hongjuan, H. Franssen, Harrie-Jan, J. A., and Harry, "State and parameter estimation of two land surface models using the ensemble kalman filter and the particle filter," Sep 2017. [Online]. Available: <https://www.hydrol-earth-syst-sci.net/21/4927/2017>
- [32] J. Rasmussen, H. Madsen, K. H. Jensen, and J. C. Refsgaard, "Data assimilation in integrated hydrological modeling using ensemble kalman filtering: evaluating the effect of ensemble size and localization on filter performance," *Hydrology and Earth System Sciences*, vol. 19, no. 7, pp. 2999–3013, 2015. [Online]. Available: <https://www.hydrol-earth-syst-sci.net/19/2999/2015/>
- [33] A. H. Weerts and G. Y. H. El Serafy, "Particle filtering and ensemble kalman filtering for state updating with hydrological conceptual rainfall-runoff models," *Water Resources Research*, vol. 42, no. 9. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005WR004093>
- [34] D. Plaza, K. R. De, and V. Pauwels, "From the ensemble kalman filter to the particle filter: a comparative study in rainfall-runoff models\*," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 14 201 – 14 206, 2011, 18th IFAC World Congress. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1474667016459082>
- [35] G. Evensen, "The ensemble kalman filter: theoretical formulation and practical implementation," *Ocean Dynamics*, vol. 53, no. 4, pp. 343–367, Nov 2003. [Online]. Available: <https://doi.org/10.1007/s10236-003-0036-9>
- [36] —, "Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics," *Journal of Geophysical Research: Oceans*, vol. 99, no. C5, pp. 10 143–10 162. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/94JC00572>
- [37] G. Burgers, P. Jan van Leeuwen, and G. Evensen, "Analysis scheme in the ensemble kalman filter," *Monthly Weather Review*, vol. 126, no. 6, pp. 1719–1724, 1998. [Online]. Available: [https://doi.org/10.1175/1520-0493\(1998\)126<1719:ASITEK>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2)
- [38] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems*. Springer, 2010.
- [39] D. Work, O.-P. Tossavainen, S. Blandin, A. Bayen, T. Iwuchukwu, and K. Tracton, "An ensemble kalman filtering approach to highway traffic estimation using gps enabled mobile devices," 01 2008, pp. 5062–5068.
- [40] A. Seppnen, M. Vauhkonen, E. Somersalo, and J. P. Kaipio, "State space models in process tomography approximation of state noise covariance," *Inverse Problems in Engineering*, vol. 9, no. 5, pp. 561–585, 2001. [Online]. Available: <https://doi.org/10.1080/174159701088027781>

- [41] J. P. KAIPIO, P. A. KARJALAINEN, E. SOMERSALO, and M. VAUHKONEN, “State estimation in time-varying electrical impedance tomography,” *Annals of the New York Academy of Sciences*, vol. 873, no. 1, pp. 430–439. [Online]. Available: <https://nyaspubs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1749-6632.1999.tb09492.x>
- [42] A. Seppnen, M. Vauhkonen, P. J Vauhkonen, E. Somersalo, and J. Kaipio, “State estimation with fluid dynamical evolution models in process tomography: Eit application,” *Inverse Problems - INVERSE PROBL*, vol. 17, 08 2000.
- [43] A. Lehtikoinen, S. Finsterle, A. Voutilainen, M. Kowalsky, and J. Kaipio, “Dynamical inversion of geophysical ert data: state estimation in the vadose zone,” *Inverse Problems in Science and Engineering*, vol. 17, no. 6, pp. 715–736, 2009. [Online]. Available: <https://doi.org/10.1080/17415970802475951>
- [44] D. Baroudi, J. Kaipio, and E. Somersalo, “Dynamical electric wire tomography: a time series approach,” *Inverse Problems*, vol. 14, no. 4, pp. 799–813, aug 1998. [Online]. Available: <https://doi.org/10.1088%2F0266-5611%2F14%2F4%2F003>
- [45] D. B. Work, S. Blandin, O.-P. Tossavainen, B. Piccoli, and A. M. Bayen, “A Traffic Model for Velocity Data Assimilation,” *Applied Mathematics Research eXpress*, vol. 2010, no. 1, pp. 1–35, 04 2010. [Online]. Available: <https://doi.org/10.1093/amrx/abq002>
- [46] “Mobile millennium.” [Online]. Available: <http://traffic.berkeley.edu/>
- [47] G. Evensen, *Data Assimilation The Ensemble Kalman Filter*. Springer Berlin, 2014.
- [48] A. J. Majda and X. T. Tong, “Performance of ensemble kalman filters in large dimensions,” *Communications on Pure and Applied Mathematics*, vol. 71, no. 5, pp. 892–937. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpa.21722>
- [49] H. H. Bauser, D. Berg, O. Klein, and K. Roth, “Inflation method for ensemble kalman filter in soil hydrology,” *Hydrology and Earth System Sciences Discussions*, pp. 1–18, 03 2018.
- [50] Y. Shi, K. J. Davis, F. Zhang, C. J. Duffy, and X. Yu, “Parameter estimation of a physically-based land surface hydrologic model using an ensemble kalman filter: A multivariate real-data experiment,” *Advances in Water Resources*, vol. 83, pp. 421 – 427, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0309170815001359>
- [51] L. S. Huning and S. A. Margulis, “Climatology of seasonal snowfall accumulation across the sierra nevada (usa): Accumulation rates, distributions, and variability,” *Water Resources Research*, vol. 53, no. 7, pp. 6033–6049. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017WR020915>
- [52] A. Arnold, D. Calvetti, and E. Somersalo, “Parameter estimation for stiff deterministic dynamical systems via ensemble kalman filter,” *Inverse Problems*, vol. 30, no. 10, p. 105008, sep 2014. [Online]. Available: <https://doi.org/10.1088%2F0266-5611%2F30%2F10%2F105008>
- [53] M. Rafiee, A. Tinka, J. Thai, and A. M. Bayen, “Combined state-parameter estimation for shallow water equations,” in *Proceedings of the 2011 American Control Conference*, June 2011, pp. 1333–1339.
- [54] C. Zoccarato, D. Ba, M. Ferronato, G. Gambolati, A. Alzraiee, and P. Teatini, “Data assimilation of surface displacements to improve geomechanical parameters of gas storage reservoirs,” *Journal of Geophysical Research: Solid Earth*, vol. 121, no. 3, pp. 1441–1461. [Online]. Available: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JB012090>
- [55] H. Steven L., L. E., Clark, and M. P., “Towards simplification of hydrologic modeling: identification of dominant processes,” Nov 2016. [Online]. Available: <https://www.hydrol-earth-syst-sci.net/20/4655/2016/>

# Appendix A

## Model Reference

### A.1 Model Parameters Reference

This section consists of Table A.1.

### A.2 Model Variables Reference

This section consists of Table A.2.

Symbol	PRMS symbol	Quantity	Type	Range	Units	Description
Pmin	ppt_rad_adj	1	monthly	0.0, 0.5	inches	minimum P to adjust solar radiation
$\phi$	dday_slope	1	monthly	0.2, 0.9	$^{\circ}\text{day}/^{\circ}\text{F}$	slope of degree-day equation
$\beta$	dday_intcp	1	monthly	-60.0, 10.0	$^{\circ}\text{day}$	intercept of degree-day equation
s	radj_sppt	1	constant	0.0, 1.0	-	adjustment factor for summer solar rad with $P > P_{\text{min}}$
w	radj_wppt	1	constant	0.0, 1.0	-	adjustment factor for winter solar rad with $P > P_{\text{min}}$
$\lambda_2$	tmin_lapse	1	monthly	-10.0, 10.0	$^{\circ}\text{F}/\text{mile}$	slope of minimum temperature interpolation
$\lambda_1$	tmax_lapse	1	monthly	-10.0, 10.0	$^{\circ}\text{F}/\text{mile}$	slope of maximum temperature interpolation
$\beta_2$	tmin_adj	per HRU	constant	-10.0, 10.0	$^{\circ}\text{F}$	physiographic adjustment to minimum T
$\beta_1$	tmax_adj	per HRU	constant	-10.0, 10.0	$^{\circ}\text{F}$	physiographic adjustment to maximum T
ds	covden_sum	per HRU	constant	0.0, 1.0	-	vegetation areal coverage during summer
dw	covden_win	per HRU	constant	0.0, 1.0	-	vegetation areal coverage during winter
Tms	tmax_allsnow	1	monthly	-10.0, 40.0	$^{\circ}\text{F}$	Tmax below which all P is snow
Tmr	tmax_allrain	1	constant	-8.0, 60.0	$^{\circ}\text{F}$	Tmax above which all P is rain
$\zeta$	adjmix_rain	1	monthly	0.6, 1.4	-	rain fraction of mixed rain-snow P
CrS	srain_intcp	per HRU	constant	0.0, 1.0	inches	canopy summer rain interception storage capacity
Crw	wrain_intcp	per HRU	constant	0.0, 1.0	inches	canopy winter rain interception storage capacity
Cs	snow_intcp	per HRU	constant	0.0, 1.0	inches	canopy winter snow interception storage capacity
jc	jh_coef	1	monthly	0.005, 0.06	$/^{\circ}\text{F}$	coef. used in ETp
$\alpha$	albedo	per HRU	constant	0.0, 1.0	-	fraction of radiation reflected by snow
$\psi$	rad_trnfc	per HRU	constant	0.0, 1.0	-	transmission coefficient for Rsw through winter canopy
$\epsilon$	emis_noppt	1	constant	0.757, 1.0	-	emissivity of air with no P
$\omega$	cecn_coef	1	monthly	2.0, 10.0	$\text{cal}/^{\circ}\text{C}$	convection & condensation coefficient
$\xi$	potet_sublim	1	constant	0.0, 1.0	-	fraction of ETp sublimated
$\rho_{\text{init}}$	den_init	1	constant	0.01, 0.5	$\text{g}/\text{cm}$	density of new snow P
$\rho_{\text{max}}$	den_max	1	constant	0.1, 0.8	$\text{g}/\text{cm}$	maximum average snowpack density
$\tau$	settle_const	1	constant	0.01, 0.5	-	settlement time constant for snowpack
fi	hru_percent_imperv	per HRU	constant	0.0, 0.999	-	HRU impervious fraction
Simax	imperv_stor_max	per HRU	constant	0.0, 0.1	inches	maximum Simp
Sszmax	soilmoist_max	per HRU	constant	0.001, 60.0	inches	maximum soilzone water holding capacity
$\alpha_1$	smidx_coef	per HRU	constant	0.001, 0.06	-	linear surface runoff (Fsr) coefficient
$\theta_1$	smidx_exp	per HRU	constant	0.1, 0.5	1/inch	exponential surface runoff coefficient
$\alpha_3$	ssrcoef_sq	per SSR	constant	0.0, 1.0	-	coefficient routing Sss to streamflow
$\beta_3$	ssrcoef_lin	per SSR	constant	0.0, 1.0	-/day	linear coefficient routing Sss to streamflow
$\alpha_2$	ssr2gw_rate	per SSR	constant	0.05, 0.8	-/day	linear coefficient routing Sss to Sgw
$\theta_2$	ssr2gw_exp	per SSR	constant	0.0, 3.0	-	exponential coefficient routing Sss to Sgw
smax	ssrmax_coef	per SSR	constant	1.0, 20.0	inches	coefficient routing Sss to Sgw
$\alpha_4$	gwflow_coef	per GWR	constant	0.001, 0.5	-/day	linear coefficient routing Sgw to streamflow
$\alpha_5$	gwsink_coef	per GWR	constant	0.0, 1.0	-/day	linear coefficient for groundwater sink
Fzgwmax	soil2gw_max	per HRU	constant	0.0, 5.0	inches	max soil excess water routed to gwStr
Acurve	snarea_curve	1	constant	0.0, 1.0	-	snow area depletion curve: fscA vs SWE <sub>max</sub>
SWE <sub>max</sub>	snarea_threah	per HRU	constant	0.0, 200.0	inches	max SWE for each HRU below which snow patch occurs
Asr	carea_max	per HRU	constant	0.0, 1.0	-	max areal fraction contributing to surface runoff Fsr

Table A.1: Model parameters

Symbol	PRMS symbol	Quantity	Units	Description
Rsw	swrad	per HRU	Langleys	shortwave radiation
r	sof	per HRU	-	ratio of actual to potential daily solar radiation
Pn	ppt_net	per HRU	inches	total net precipitation
Pt	throughfall	per HRU	inches	precipitation throughfall after interception
Sc	AvailCanSt	per HRU	acre-inch	available storage in canopy
ETp	potet	per HRU	inches	potential evapotranspiration
ETpa	apet	per HRU	inches	available potential evapotranspiration
$\Delta E$	tcal	per HRU	Langleys	net snowpack energy balance
I	cal	per HRU	Langleys	total incoming longwave
Qv	ceesub	per HRU	Langleys	convection and latent heat from condensation
Rn	swn	per HRU	Langleys	net shortwave radiation
Ip	lwp	per HRU	Langleys	perfect black-body emission
Tavg	temp	per HRU	$^{\circ}\text{F}$	average temperature
Qc	qcond	per HRU	Langleys	conducted heat
B	sub	per HRU	inches	sublimation
Pns	net_snow	per HRU	inches	snow portion of Pn
D	pk_depth	per HRU	inches	snowdepth
M	snowmelt	per HRU	inches	snowmelt
fsa	snowcov_area	perHRU	.	fractional snow covered area
Pnr	net_rain	per HRU	inches	rain portion of Pn
Psri	hru_sroffi	per HRU	inches	impervious surface runoff
ei	hru_impervevap	per HRU	inches	imprevious region evaporation

Table A.2: Model variables

## **Appendix B**

# **Additional Simulation Years**

### **B.1 Water Year 2011 Results**

#### **B.1.1 SWE update**

This section consists of Figure B.1.

#### **B.1.2 Joint update**

This section consists of Figures B.2 and B.3.

### **B.2 Water Year 2014 Results**

#### **B.2.1 SWE update**

This section consists of Figure B.4.

#### **B.2.2 Joint update**

This section consists of Figures B.5 and B.6.

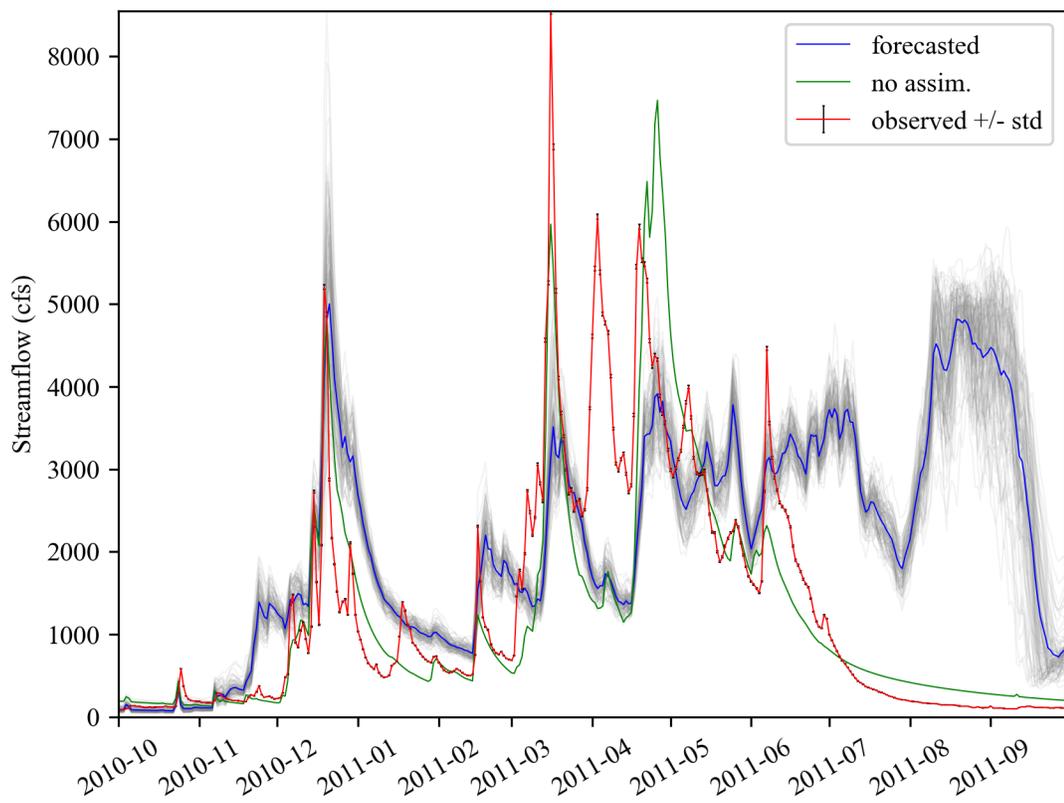


Figure B.1: Water year 2011 basin average streamflow (SWE update).

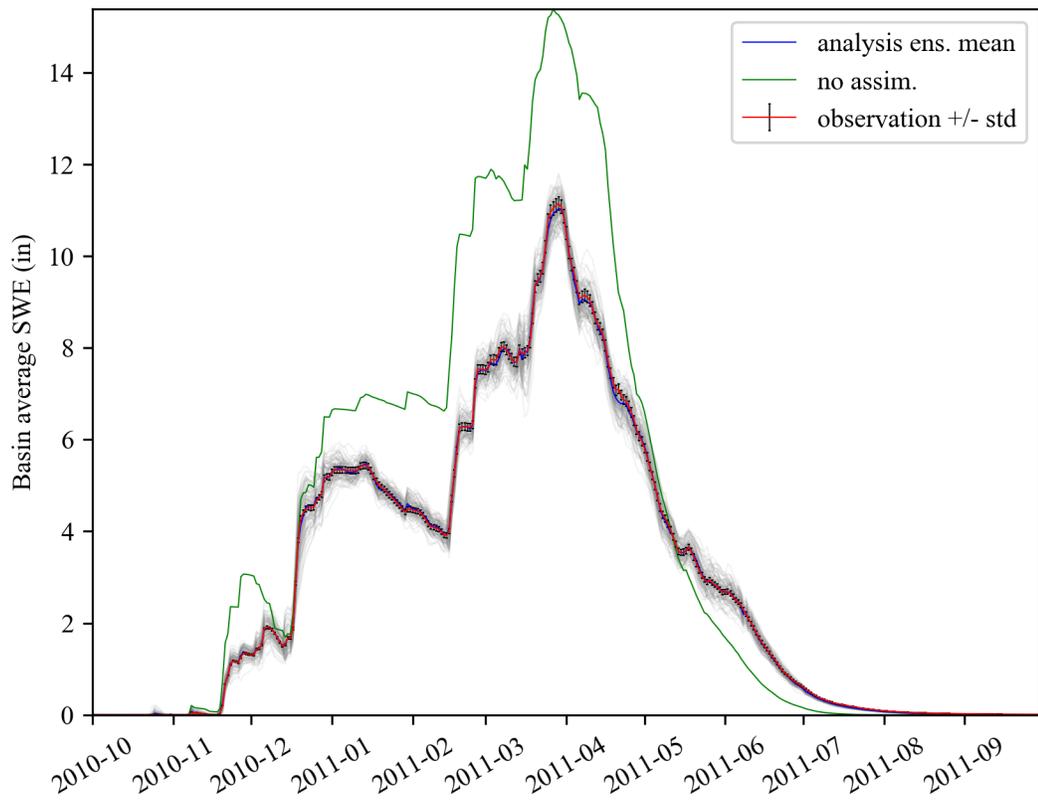


Figure B.2: Water year 2011 basin average SWE. (SWE or Joint update)

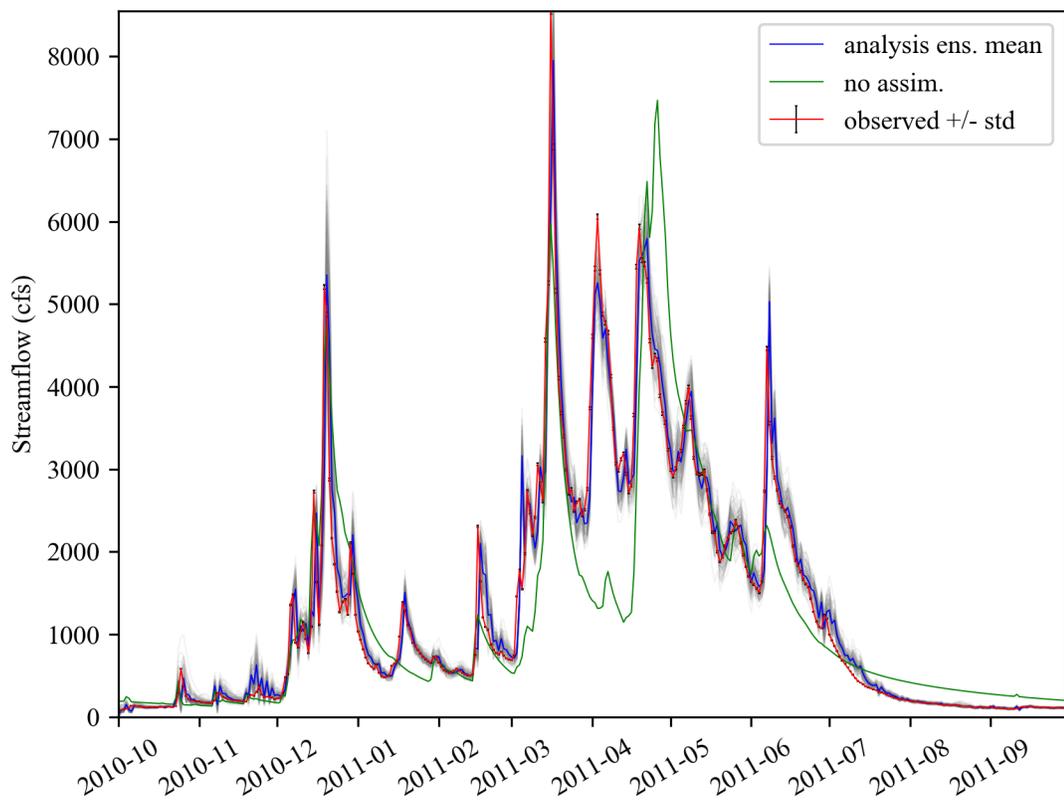


Figure B.3: Water year 2011 basin average streamflow (Joint update).

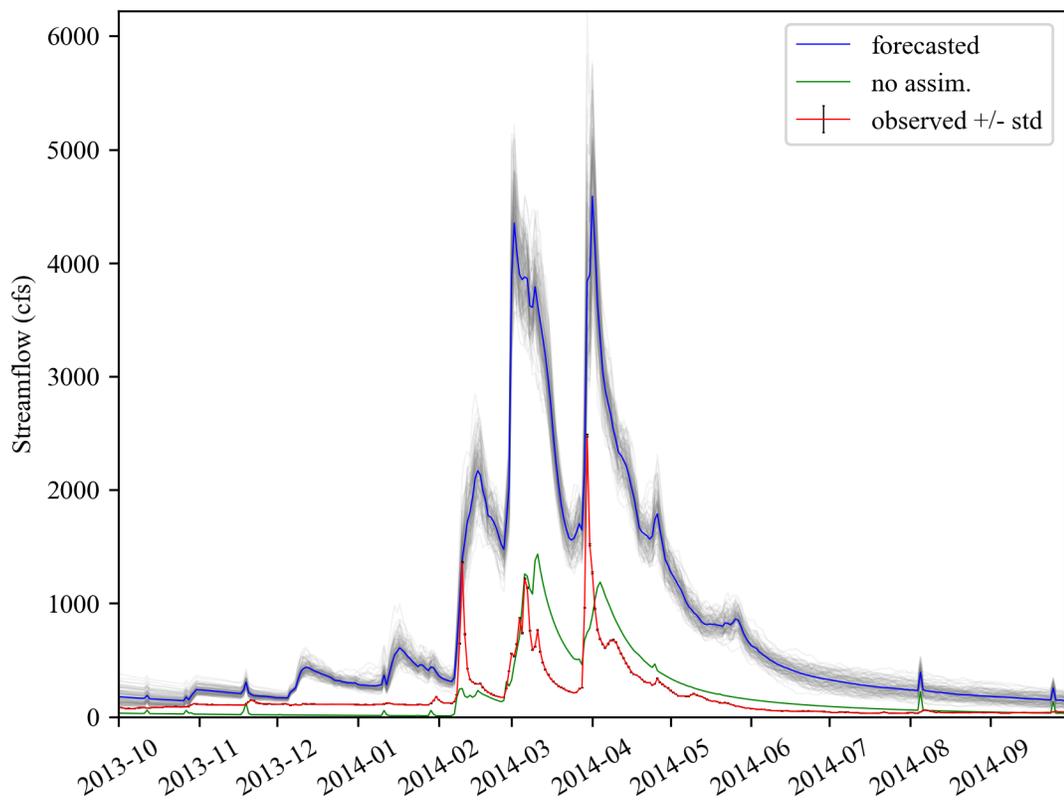


Figure B.4: Water year 2014 basin average streamflow (SWE update).

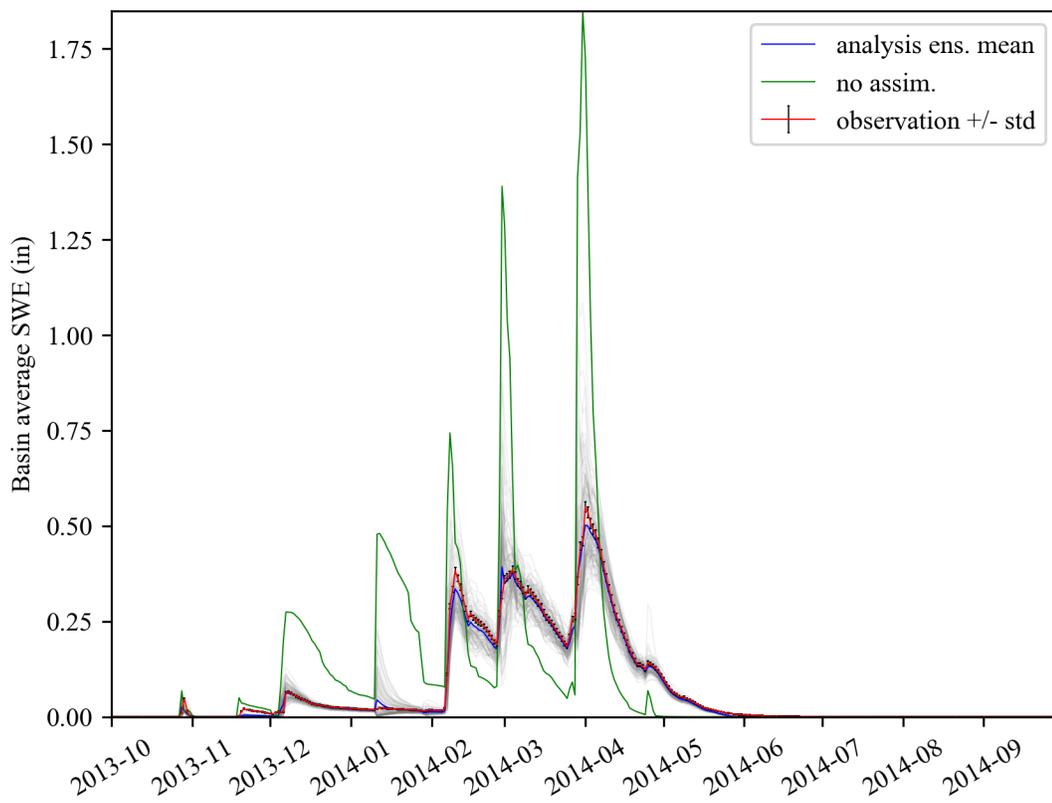


Figure B.5: Water year 2014 basin average SWE. (SWE or Joint update)

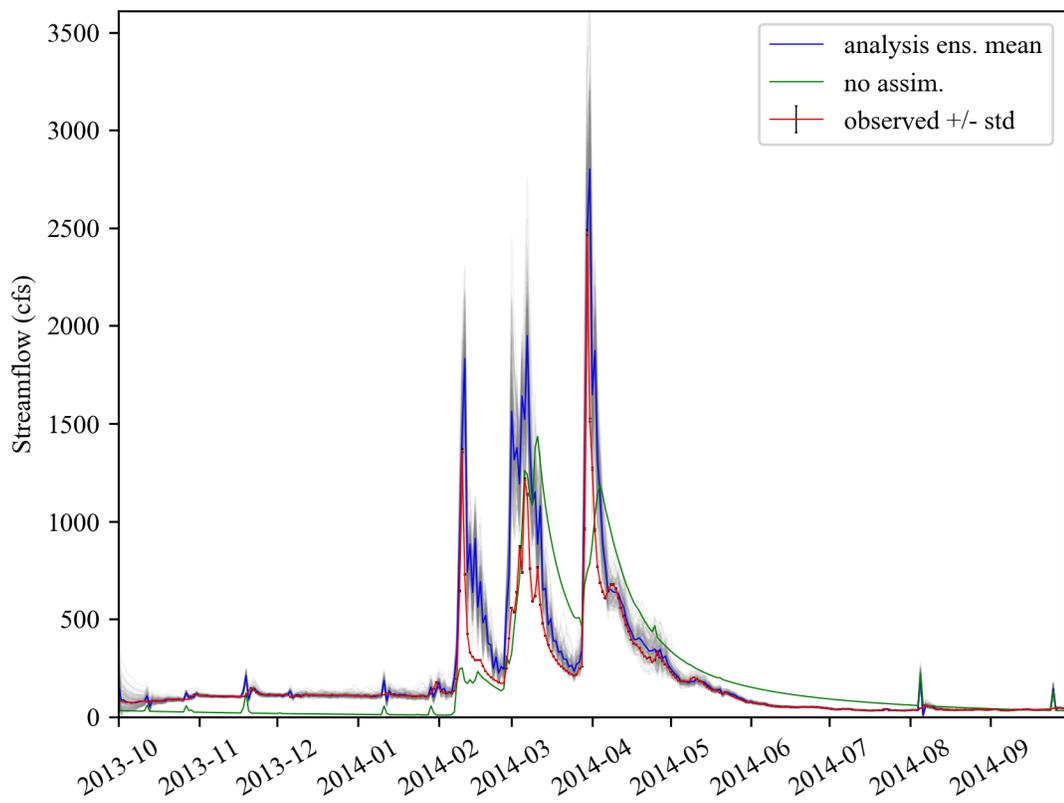


Figure B.6: Water year 2014 basin average streamflow. (Joint update)