

APPLICATION OF AN ENSEMBLE KALMAN FILTER TO A SEMI-DISTRIBUTED HYDROLOGICAL FLOOD FORECASTING SYSTEM IN ALPINE CATCHMENTS

Alain, Foehn¹

Plateforme de Constructions Hydrauliques, Faculté de l'Environnement Naturel, Architectural et Construit (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Suisse
alain.foehn@alumni.epfl.ch

Anne, Schwob

Plateforme de Constructions Hydrauliques, Faculté de l'Environnement Naturel, Architectural et Construit (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Suisse
anne.schwob@epfl.ch

Damiano, Pasetto

Laboratoire d'écohydrologie, Faculté de l'Environnement Naturel, Architectural et Construit (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Suisse
damiano.pasetto@alumni.epfl.ch

Javier, García Hernández

Centre de recherche sur l'environnement alpin (CREALP), Sion, Suisse
javier.garcia@crealp.vs.ch

Giovanni, De Cesare

Plateforme de Constructions Hydrauliques, Faculté de l'Environnement Naturel, Architectural et Construit (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Suisse
givoanni.decesare@epfl.ch

KEY WORDS

Data assimilation, hydrological modelling, data uncertainty, data perturbation

ABSTRACT

One of the key success factors for hydrological forecasts is initial conditions that represent well the conditions of the simulated basin at the beginning of the forecast. Real-time Data Assimilation (DA) has been shown to allow improving these initial conditions. In this article, two DA approaches are compared with the reference scenario working without DA (Control). In both approaches, discharge data at gauging stations are assimilated. In the first approach, a volume-based update (VBU) compares the simulated and observed volumes over the past 24 hours before the start of a forecast to compute a correction factor used to update the initial soil water saturation in the upstream part of the semi-distributed hydrological model. In the second approach, an ensemble Kalman filter (EnKF) is implemented to account for the uncertainty in initial conditions, precipitation, temperature and discharge data. The comparison is carried out over two sub-basins of the Upper Rhone River basin upstream of Lake Geneva, in Switzerland, where the MINERVE flood forecasting and management system is implemented. Results differ over the two studied basins. In one basin, the two DA simulations perform better than the Control simulation, with the EnKF simulation providing the best forecasting performance. In the second basin, where the Control simulation performs best, possible challenges with hydropower-based discharges are highlighted.

¹ Corresponding author

1. INTRODUCTION

Severe flooding events in recent decades have increased the need for reliable forecasting systems in Alpine catchments. Following the major flood of October 2000 in the Canton of Valais (Switzerland), an operational forecasting system (MINERVE) has been set up in 2013 and provides a tool for decision-making tasks [Jordan2007, García2011, García2014].

Despite efforts on the calibration of the model, hydrological forecasts are subject to a number of uncertainties [Srikanthan2008]. First, meteorological inputs suffer from uncertainties both on the measurements and the meteorological forecast. This is true for the precipitation, but also for the temperature that can be difficult to estimate accurately in alpine catchments. Second, the model represents a simplification of the real system which implies errors in the simulated discharges. Third, the discharge measurements also suffer from uncertainties. All these uncertainties must be considered when analyzing the performances of a hydrological forecasting system.

Data assimilation (DA) techniques are mathematical tools developed to correct the model results using available system observations and taking into account explicitly the different sources of uncertainty. One of the first developed techniques was the Kalman filter (KF) [Kalman1960], largely used for linear models. To allow working with non-linear models, alternative methods have been developed. One of these variants is the ensemble Kalman filter (EnKF), a Monte Carlo approach of the KF, performing an ensemble of model runs [Evensen1994]. Weerts et al. 2006 compared the EnKF with Particle filtering (another DA approach) and showed that EnKF was more robust and outperformed the two other analyzed filters. In the following study, an implementation of the EnKF with updating of state variables is explored using the SOCONT and GSM semi-distributed conceptual models. The performance gain is evaluated over two flood events concerning two different rivers located within the Upper Rhone River basin in Switzerland.

2. MATERIAL AND METHODS

2.1 Watersheds and data

The analysis concerns the streamflow simulation at the outlet of the Aigle and Reckingen sub-catchments of the Upper Rhone River basin. In each sub-catchment, a period of high flow is analyzed over which weather radar data is available. The evaluated high flow periods occurred in 2012 in Reckingen and in 2015 in Aigle. Corresponding characteristics of the sub-catchments are given in **Table 1**.

Station	River	Area [km ²]	Catchment mean elevation [m a.s.l.]	Glaciation [%]	Hourly peak flow [m ³ /s]	Return period of the studied event [years]
Reckingen	Rhone	214	2305	11.8	154.8	36
Aigle	Grande-Eau	132	1562	0.8	60.1	10

Table 1: Main characteristics of the catchments

Both rivers are equipped with a small run-of-river hydropower plant. The one in the Rhone (Reckingen basin) has an equipped discharge of 5.7 m³/s (yearly average discharge in 2012: 10.6 m³/s). The equipped discharge over the Grande-Eau was 2.5 m³/s until 2015 and is 6.5 m³/s since April 2016 (yearly average discharge in 2015: 4.02 m³/s). In addition, the Grande-Eau catchment receives water diverted from the Lac d'Arnon, a reservoir located in a nearby catchment with a capacity of 11 million m³. When the natural discharge in the Grande-Eau River is below the installed capacity of the two successive run-of-river hydropower plants, the water from the Lac d'Arnon is turbed at a third more upstream hydropower plant with a capacity of 1.75 m³/s. The operating data of the scheme were not available for the present study.

Precipitation data are taken from a spatial product of precipitation following the regression co-kriging approach presented in Foehn2018, in which radar data are combined with two networks of ground station data into a 1 km x 1 km gridded data with a one hour temporal resolution. Temperature data are interpolated from the values observed at the meteorological stations with a constant vertical lapse rate of -5.5°C/1000 m.

In order to evaluate the performance of the data assimilation method, observed data of precipitation and temperature are used for both the assimilation and forecast simulations; i.e. no forecast data are considered in the present analysis. This is mainly due to the limited temporal coverage of historical weather forecast data. Furthermore, working with observed values can be seen as introducing less uncertainty as weather forecasts. Nevertheless, it must be mentioned that the Aigle and Reckingen basins are neither well covered by the SwissMetNet automatic monitoring network of the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss) nor by the Swiss weather radar network, due to the complex topography of the Upper Rhone River basin. Therefore, the radar data are locally subject to possible underestimations.

To account for these underestimations, the spatial product of precipitation was multiplied by a correction factor over both basins. For Reckingen, a factor of 1.6 has been applied. This value is probably too low according to the difference between simulated and observed flow, but higher values were not investigated in this study. For the Grande-Eau, spatially varying values between 1.15 and 1.4 were considered. Variations in the factor are justified by the varying visibility of the weather radar data over the basin, with the Wildhorn peak (3250 m a.s.l.) considerably reducing the visibility over the basin of the closest weather radar (at Pointe de la Plaine Morte), located at only 2926 m a.s.l.

2.2 GSM-SOCONT model

Simulations are performed with the rainfall-runoff semi-distributed conceptual GSM-SOCONT (Glacier and SnowMelt – Soil CONTRibution) model [Schaepli2005, GarcíaHernández2019], illustrated in Figure 1. The main parameters of the model are given in Table 2.

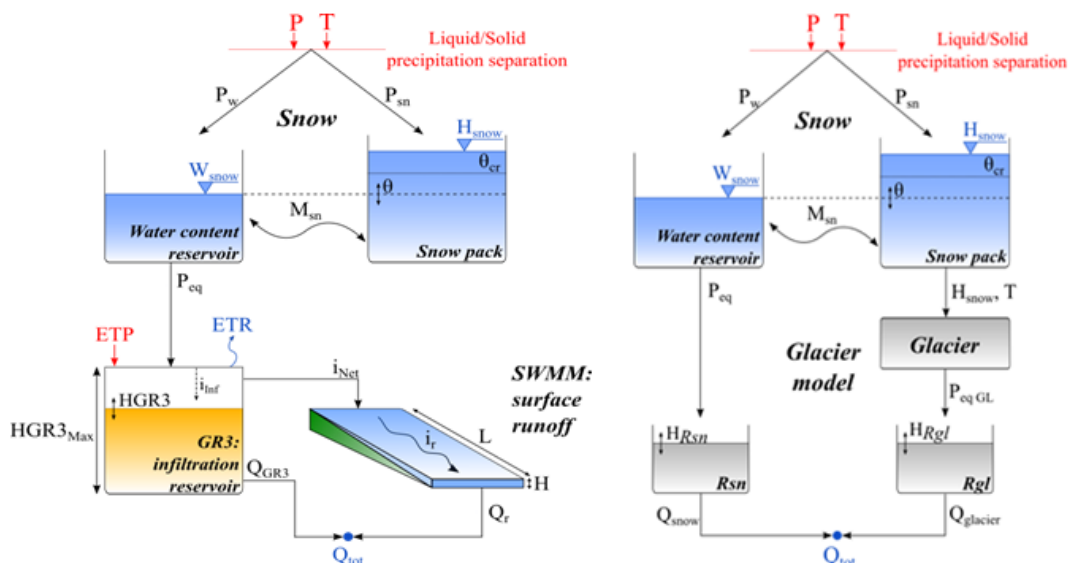


Figure 1: Schematic of the SOCONT (left) and GSM (right) rainfall-runoff models ([GarcíaHernández2019])

For the semi-distributed model, the studied catchment is divided into elevation bands with an elevation range not exceeding 500 m. This allows accounting for the vertical evolution of temperature. The GSM model is used for glacial elevation bands and the SOCONT model for non-glacial elevation bands. For both models, the snowmelt rate is controlled by the snowmelt coefficient. For the GSM model, when the surface is free of snow, glacial melt is considered based on the glacier melt coefficient and the temperature. For the SOCONT model, the generated discharge at the outlet of the basin is controlled by the absorption capacity of the soil, the release coefficient of the infiltration reservoir as well as the roughness of the surface runoff surface. The model has been calibrated using the same spatial products of precipitation and temperature used for the assimilation.

Parameters	Units	Model SOCONT
Asn	mm/d/°C	Reference degree-day snowmelt coefficient
HGR3 _{Max}	m	Maximum height of infiltration reservoir
KGR3	1/s	Release coefficient of infiltration reservoir
Kr	m ^{1/3} /s	Strickler coefficient of surface runoff
Variable	Units	Model GSM
Asn	mm/d/°C	Reference degree-day snowmelt coefficient
Agl	mm/d/°C	Reference degree-day glacier melt coefficient
Ksn	1/d	Release coefficient of snow melt reservoir
Kgl	1/d	Release coefficient of glacier melt reservoir

Table 2: Main parameters of models SOCONT and GSM models

2.3 Data assimilation techniques

Two DA techniques are explored in this work to improve the values of the system state variables at the beginning of each forecast. The first method is based only on a volume comparison between simulated and observed discharges over the past 24h. This method has been used operationally for three years in the MINERVE system. The second method consists in an ensemble Kalman filter, where an ensemble of possible model trajectories is corrected by assimilating the current available discharge observation.

The objective of the data assimilation is reducing the discrepancy between the simulated state variables and the actual state of the system to improve the forecast performance of the model. To achieve this, the state variables of the model are updated using the available discharge observations.

In the SOCONT model, the state variables which initial conditions need to be updated are the water level in the soil reservoir (HGR3), the surface runoff water level (Hr) and the Snow water equivalent (SWE) which are evaluated at each spatial node of the model. Furthermore, the snow water equivalent (SWE) as well as the water level in the glacial melt reservoir and the snow melt reservoir from the GSM model could be updated but this was not explored in the study. In the following, \mathbf{x}_t indicates the system state vector of dimension $nstate=3 \times nnodes$, which elements are the SOCONT state variables HGR3, Hr and SWE in all modeled nodes. The temporal evolution of the state vector obtained with GSM-SOCONT is formally represented by the following dynamical model:

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) \quad (1)$$

where the function f is the set of equations linking the state of the system from time t to $t+1$, \mathbf{u}_t^i represents the vector of model inputs (here, spatial maps of temperature and precipitation) and \mathbf{w}_t is the possible occurrence of system noise (not directly considered in this analysis).

The assimilation procedure is performed using real measurements of the system given by discharge data based on water level observations at the outlet of the watershed, provided by the Swiss Federal Office for the Environment (FOEN). Therefore, the initial discharge of the (kinematic wave) river reaches (Qini) of the model is also considered in the state vector. Observations are indicated with the vector \mathbf{y}_t of dimension $nobs$ (here $nobs=1$, since the observation is the discharge at time t at a unique gauging station). The link between the state variables \mathbf{x}_t and the observations, \mathbf{y}_t , is provided by the following observation operator:

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{v}_t \quad (2)$$

where \mathbf{H} is a projection matrix (dimension $nobs \times nstate$), and \mathbf{v}_t represents possible measurement errors (see Section 3.1.3). The matrix \mathbf{H} is constructed with a value of 1 in correspondence of the state variable being observed, and 0 elsewhere [Clark2008].

In the following, the three assimilation procedures considered to produce the forecasts of discharge are described.

2.3.1 Control simulation (Control)

The reference scenario is computed by running the model (eq. 1) without any perturbation of input data (i.e., \mathbf{u}_t corresponds to the nominal values of precipitation and temperature) and without considering the discharge measurements during the simulation (no data assimilation). This simulation corresponds to what is frequently called the open-loop scenario.

2.3.2 Volume-based update (VBU)

The first data assimilation approach is the volume-based update (VBU), which is based on the comparison between the volumes of the observed and the simulated discharges. The only updated state variable is the soil water content (HGR3) of the SOCONT models. The value of the corresponding HGR3 at the start of the VBU simulation (24 hours before the beginning of the hydrological forecast) is iteratively changed (up to ten iterations) so that the simulated volumes throughout the simulation are as close as possible to the observed ones. The implemented approach limits the saturation at the start of the VBU simulation to 75% of the maximum water content in the soil, in order to avoid a too much reactive result. Note that the VBU simulation does not take into consideration the uncertainties in the inputs, model, or observations. The update considers all the discharge observations collected during the 24 hours before the forecast. Similarly to the control simulation, VBU provides as output only one model trajectory.

2.3.3 Ensemble Kalman filter (EnKF)

The second data assimilation approach is an implementation of the ensemble Kalman filter (EnKF), which better allows the update of state variables taking into consideration model and measurements uncertainties. The EnKF is an adaptation of the Kalman filter in which the covariance matrix is replaced by the sample covariance computed from an ensemble of possible state vectors. The EnKF is based on the sequential repetition of two steps: the prediction step and the analysis (or assimilation) step. In the prediction step the different members of the ensemble are independently advanced in time by running the model (eq. 1) applying different random samples of uncertain forcing terms [Srikanthan2008, Abaza2015]:

$$\mathbf{x}_{t+1}^{i,p} = f(\mathbf{x}_t^{i,a}, \mathbf{u}_t^i, \mathbf{w}_t), \quad i = 1, \dots, nens \quad (3)$$

where $\mathbf{x}_{t+1}^{i,p}$ is the i^{th} predicted ensemble state at time $t+1$, $\mathbf{x}_t^{i,a}$ is the i^{th} assimilated ensemble state at time t , \mathbf{u}_t^i represents the possible occurrence of model and/or input uncertainties (here, a perturbation of temperature and precipitation, see Section 3.1). The index “p” indicates the prediction and the index “a” the analysis.

In the analysis step, the predicted state variables $\mathbf{x}_{t+1}^{i,p}$ are updated using the newly available observation, \mathbf{y}_{t+1} . The $nens$ ensemble members are combined into the $nstate \times nens$ model state matrix, that is:

$$\mathbf{X}_{t+1}^p = (\mathbf{x}_{t+1}^{1,p}, \mathbf{x}_{t+1}^{2,p}, \dots, \mathbf{x}_{t+1}^{nens,p}) \quad (4)$$

where $nstate$ is the number of state variables and $nens$ is the number of ensemble members. The ensemble mean is given by:

$$\bar{\mathbf{x}}_{t+1}^p = \frac{1}{nens} \sum_{i=1}^{nens} \mathbf{x}_{t+1}^{i,p} \quad (5)$$

which is used to compute the model error for each ensemble member i :

$$\mathbf{E}_{t+1}^p = (\mathbf{x}_{t+1}^{1,p} - \bar{\mathbf{x}}_{t+1}^p, \mathbf{x}_{t+1}^{2,p} - \bar{\mathbf{x}}_{t+1}^p, \dots, \mathbf{x}_{t+1}^{nens,p} - \bar{\mathbf{x}}_{t+1}^p) \quad (6)$$

where \mathbf{E}_{t+1}^p is the ensemble anomaly.

The ensemble model covariance matrix ($nstate \times nstate$) can then be defined as follows:

$$\mathbf{P}_{t+1}^p = \frac{1}{N-1} \mathbf{E}_{t+1}^p \mathbf{E}_{t+1}^{pT} \quad (7)$$

In order for the EnKF to maintain sufficient spread in the ensemble and prevent filter divergence [Burgers1998], observations are perturbed in accordance with the measurement error (eq. 2) to create a $nobs \times nens$ vector of observations \mathbf{Y}_{t+1} . The analysis equation is then given by the EnKF update, which is optimal in the case of errors with a Gaussian distribution:

$$\mathbf{X}_{t+1}^a = \mathbf{X}_{t+1}^p + \mathbf{K}_{t+1}(\mathbf{Y}_{t+1} - \mathbf{H}_{t+1}\mathbf{X}_{t+1}^p) \quad (8)$$

where the matrix \mathbf{K}_{t+1} ($nstate \times nobs$) is called the Kalman gain:

$$\mathbf{K}_{t+1} = \mathbf{P}_{t+1}^p \mathbf{H}_{t+1}^T (\mathbf{H}_{t+1} \mathbf{P}_{t+1}^p \mathbf{H}_{t+1}^T + \mathbf{R}_{t+1})^{-1} \quad (9)$$

and where \mathbf{R}_{t+1} is the $nobs \times nobs$ observation error covariance matrix.

Note that each ensemble member is updated separately. Moreover, the forecast computed after an EnKF update consists of an ensemble of model trajectories, from which it is possible to quantify the uncertainty associated to the forecast. In the EnKF implementation realized in this study, a new assimilation is performed every two hours. Furthermore, the constraint of maximum 75% of soil saturation used in the VBU is not applied in the implementation of the EnKF. This choice was motivated by the higher sophistication of the EnKF approach using the ensemble approach. Nevertheless, EnKF can also result in high oversaturation in case of large differences between simulated and observed discharges.

3. EXPERIMENTAL SET-UP

3.1 Uncertainties in input and output

Model uncertainties are quantified by the ensemble Kalman filter through the empirical probabilistic distribution of the ensemble members. Input forcing data (precipitation and temperature in this study) are perturbed to provide each member a different input and thereby ensure spread in the ensemble. Members are initialized with a perturbation of the state variables of the model.

3.1.1 Precipitation uncertainty

Precipitation is perturbed following a lognormal distribution with a temporal correlation. During the prediction step of EnKF, the nominal value of precipitation is multiplied by a coefficient as follows:

$$P_t^i = P_t \times e_t^i \quad (10a)$$

with:

$$e_t^i = \exp(z_t^i) \sim \log N(m_e, \sigma_e) \quad (10b)$$

where P_t is the measured precipitation at time t , e_t^i is the multiplier coefficient for the i^{th} member at time t and P_t^i is the perturbed precipitation for the i^{th} member at time t .

At time $t=0$, z_0^i is sampled from:

$$z_0^i \sim N(0, \theta) \quad (11a)$$

with a mean equal to 0 so that the median of $\exp(z_0^i)$ equals 1 and the standard deviation θ fixed to 0.3 for Grande-Eau and 0.5 for Reckingen (0.5 corresponds to perturbation factors e_t^i with quantiles 5, 25, 50, 75 and 95% of respectively 0.44, 0.71, 1.00, 1.40 and 2.18).

In order to ensure a temporal correlation of the perturbation for a given member at time $t>0$, the time evolution of precipitation errors is simulated as follows:

$$z_t^i = \rho z_{t-1}^i + \sqrt{(1 - \rho^2)} \omega_t^i \quad (11b)$$

$$\omega_t^i \sim N(0, \theta) \quad (11c)$$

$$\rho = 1 - \frac{\Delta T}{\tau} \quad (11d)$$

where ω_t^i is the sample white noise, ρ is the temporal persistence parameter, ΔT is the simulation time step and τ the decorrelation time step. In the present study, ΔT is one hour and τ is fixed to 24 hours (following Clark2008).

In addition to the temporal correlation, the spatial correlation of the perturbation is considered by modifying for a given member all pixels with the same perturbation.

3.1.2 Temperature uncertainty

Temperature is perturbed using an additive term following a normal distribution:

$$T_t^i = T_t + s_t^i \quad (12a)$$

where T_t is the measured temperature at time t , s_t^i is the additive coefficient for the i^{th} member at time t and T_t^i is the perturbed temperature for the i^{th} member at time t .

At time $t=0$, the perturbation of the i^{th} member is given by:

$$s_0^i \sim N(0, \sigma_s^2) \quad (12b)$$

where the standard deviation σ_s is fixed to 2°C.

At time $t>0$, a temporal correlation of the perturbation is considered following Eq. (11b) to (11d) using the normal distribution given in Equation 12b and a decorrelation time fixed to 12 hours. Similarly to the precipitation, a spatially constant correction is considered for each member.

3.1.3 Measurement uncertainty

Errors in the streamflow measurements can result from both errors in the level measure and uncertainties in the rating curve used to transform water level into discharge data. The perturbed discharges are computed with an additive term as following:

$$y_t^i = y_t + \beta_t^i - \gamma \quad (13a)$$

$$\beta_t^i \sim N(0, \sigma_\beta^2) \quad (13b)$$

$$\sigma_\beta = \varepsilon_y \times y_t \quad (13c)$$

where ε_y is a hyper-parameter allowing to define the perturbation proportionally to the discharge values and γ corresponds to a correction for possible external discharge contributions to the basin, like in the case of the Grande-Eau (1.75 m³/s from the Lac d'Arnon located outside of the basin). The γ correction is randomly considered by using a uniform random variable (i.e. discharge values for 50% of the members are reduced by the defined external discharge, 50% are not modified).

3.1.4 Members initialization

Before the first assimilation, initial conditions of the state variables are also perturbed. The level in the infiltration reservoir, the level of surface runoff and the snow water equivalent height of the SOCONT models are perturbed using a normal distribution with mean 0 and a standard deviation corresponding to 40% of the original state variable value.

The original initial values are computed with a one year warm-up simulation using the same precipitation and temperature data as the one used for the data assimilation.

3.2 Performance evaluation

Two precipitation events are analyzed and the VBU and EnKF are used to assimilate streamflow measurements every 2 hours. After each assimilation, a new streamflow forecast of 3 days is produced using observed precipitation and temperature. The performances of the control scenario, VBU and EnKF are evaluated by computing the root mean squared error (RMSE) between the forecasted and measured streamflow values at different lead times L (each hour) during the forecast as followed:

$$RMSE_L = \sqrt{\frac{\sum_{j=1}^n \{Q_{L,j}^f - Q_{L,j}^{obs}\}^2}{n}} \quad (12)$$

where n is the number of forecasts produced during the single precipitation event.

3.3 Code implementation

The methodology has been implemented in the R language and environment (R Core Team_2018), in particular with the packages *parallel* for parallel computation (R Core Team 2018), *ggplot2* for the plots (Wickham 2016) and *httr* for the data acquisition (Wickham 2018).

4. RESULTS AND DISCUSSION

Simulations have been run for the two events presented in **Table 1**.

Figure 2 shows the forecasted flow about 24 hours before the peak flow for the high flow event at station Reckingen (Rhône River). For the EnKF approach, lines corresponding to all members as well as the median (red) and the mean (brown) are shown. Over the assimilation period, the difference between the VBU (green line) and the Control (blue line) are well visible, with the VBU line being closer to the observations. The red line, existing only over the two hours preceding the assimilation time (corresponding to the prediction phase of the members) is close to the observed discharge value of the assimilation time (vertical dashed line). The daily variation due to snow and ice melt is well visible in the observed discharge.

The simulated discharge, in particular for the Control simulation, is well below the observed discharge, even though the precipitation has been multiplied by 1.6. This confirms the need of modifying the

precipitation input to well reproduce the event and highlights the importance of representative input data to the model.

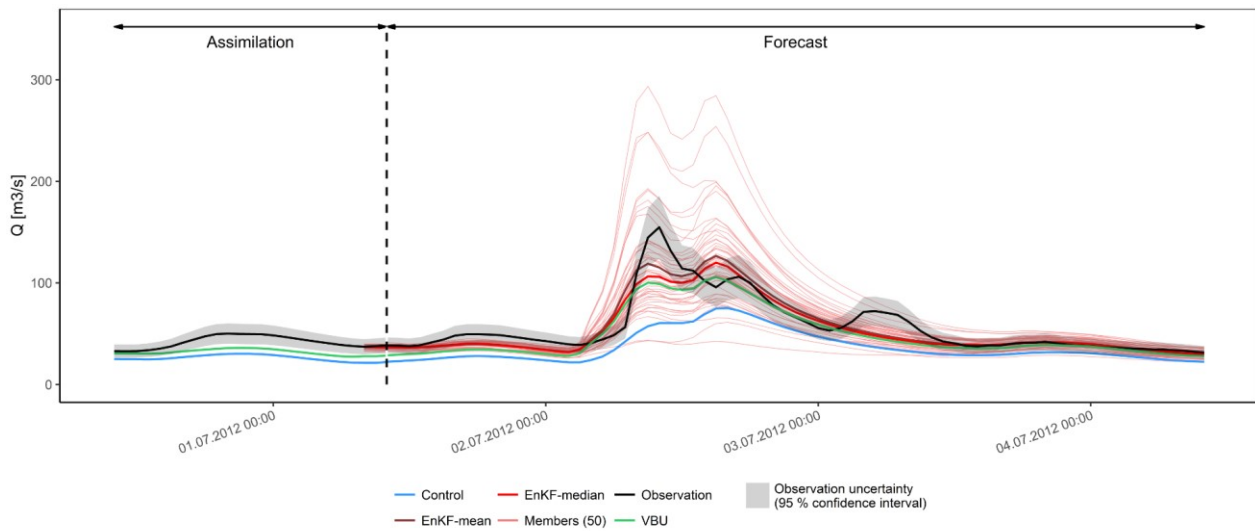


Figure 2: Forecasted discharges for the three different approaches at the Reckingen station on the Rhone River. The dashed line corresponds to 2012-07-01T10+01:00.

The RMSE values obtained during the forecast for the event are presented in Figure 3. The EnKF provides the best performance, followed by the VBU approach. The difference between the two DA approaches is not high, except over the few first forecast hours, where the EnKF results in a lower error.

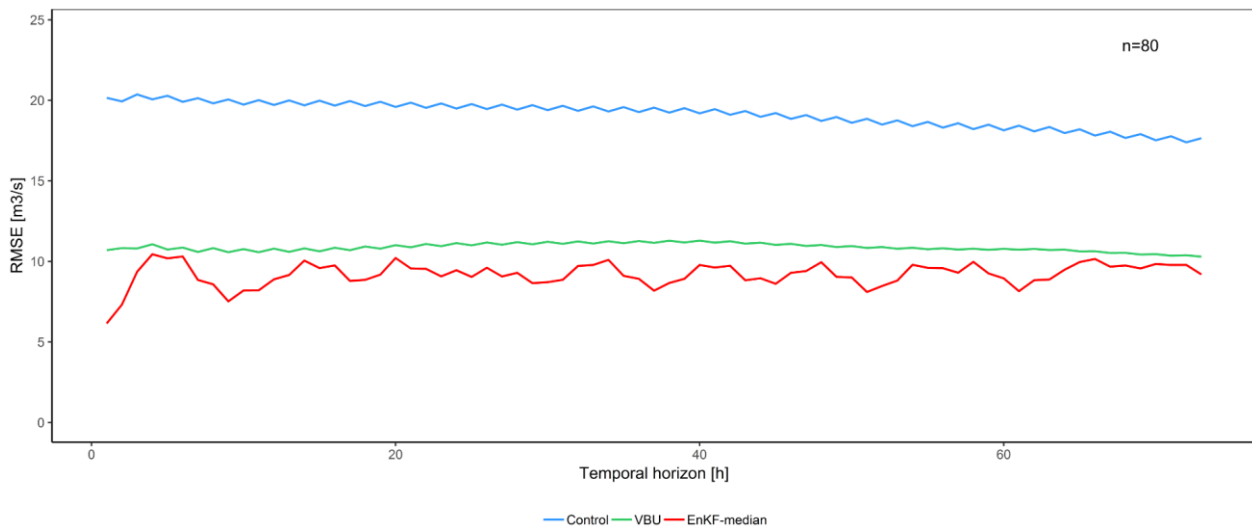


Figure 3: RMSE values obtained for the three approaches for the high flow event at the Reckingen station on the Rhone River. The n value indicates the number of runs considered in the computation of the RMSE value of each temporal horizon.

It must be mentioned that the considerable gain of the two data assimilation approaches (VBU and EnKF) is here partially an artefact. In fact, the difference between simulated and observed discharges probably rather comes, at least partially, from an underestimation in the snow and ice melt, variables which are currently not updated in the presented system. To compensate this underestimation, the VBU and EnKF approaches increase the soil water content of the non-glacial part of the basin to increase the base flow. This results during the event in a much higher peak flow as compared to the Control simulation. The hypothesis of an overcompensation of the soil reservoir is also supported by the very rapid hydrological answer of the system at the beginning of the event, suggesting a highly saturated soil. This analysis supports the need to improve the implemented EnKF approach to update also the initial level of glacial and ice melt reservoirs.

Furthermore, regarding the difference between EnKF and VBU, it probably results from the constraint used in VBU to limit the soil water storage to 75% of its maximum capacity, constraint not applied to EnKF.

Over the second studied high flow event, concerning the Grande-Eau basin, results are quite different. During the event, both the VBU and the EnKF approaches perform similar to the Control simulation (Figure 4). However, when looking at the performance over the entire event (i.e. including the forecasts generated before the beginning of the event), the Control simulation performs best (Figure 5).

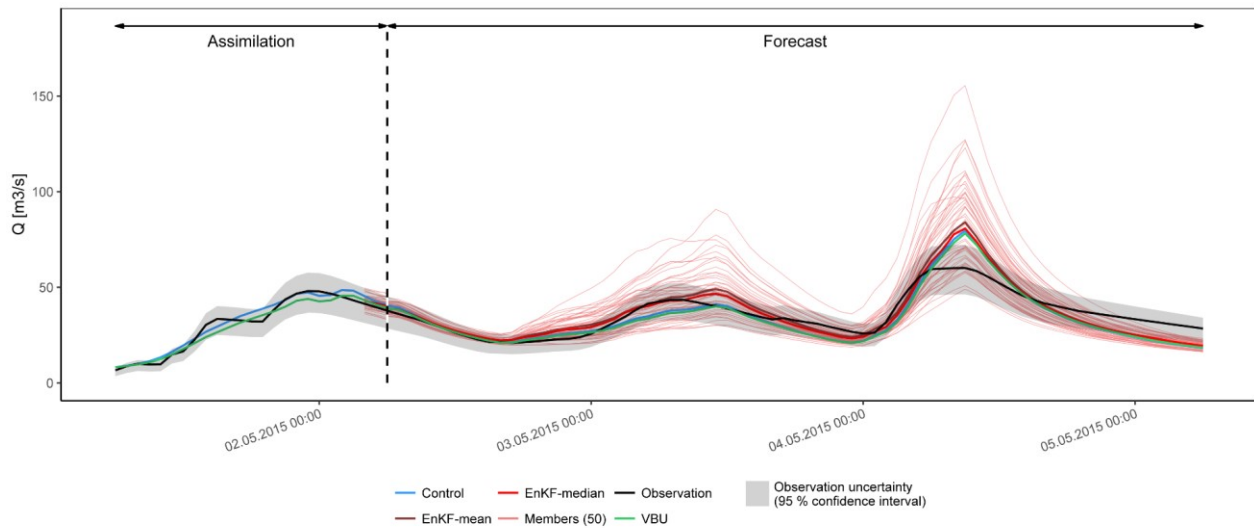


Figure 4: Forecasted discharges for the three different approaches at the Aigle station on the Grande-Eau River. The dashed line corresponds to 2015-05-02T06+01:00.

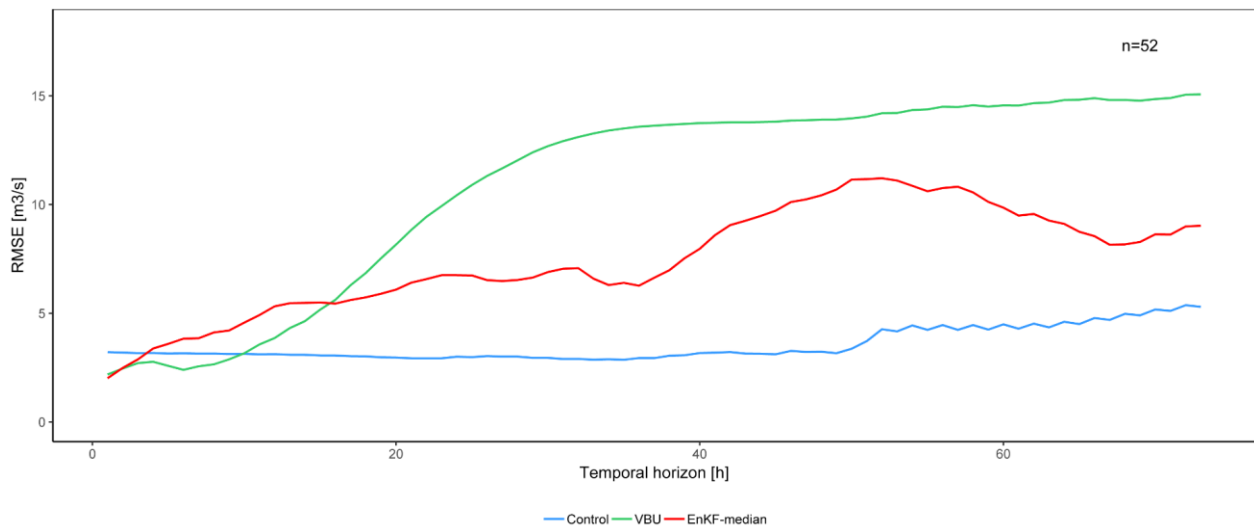


Figure 5: RMSE values obtained for the three approaches for the high flow event at the Aigle station on the Grande-Eau River. The n value indicates the number of runs considered in the computation of the RMSE value of each temporal horizon.

The better performance of the Control simulation results from the fact that when applying the data assimilation methods on the base flow, both methods fail in well forecasting the observed discharge by highly overestimating the discharge (Figure 6). This low performance of both DA approaches is here mainly linked to the external water (from the Lac d’Arnon) turbiné in the upstream part of the basin. Exclusively operated during low flows, the relative impact on the base flow is considerable. Over the studied event, the external input results in an increase of the observed discharge from $6.2 \text{ m}^3/\text{s}$ to $9.8 \text{ m}^3/\text{s}$, which represents an artificial increase of the natural discharge of 58%.

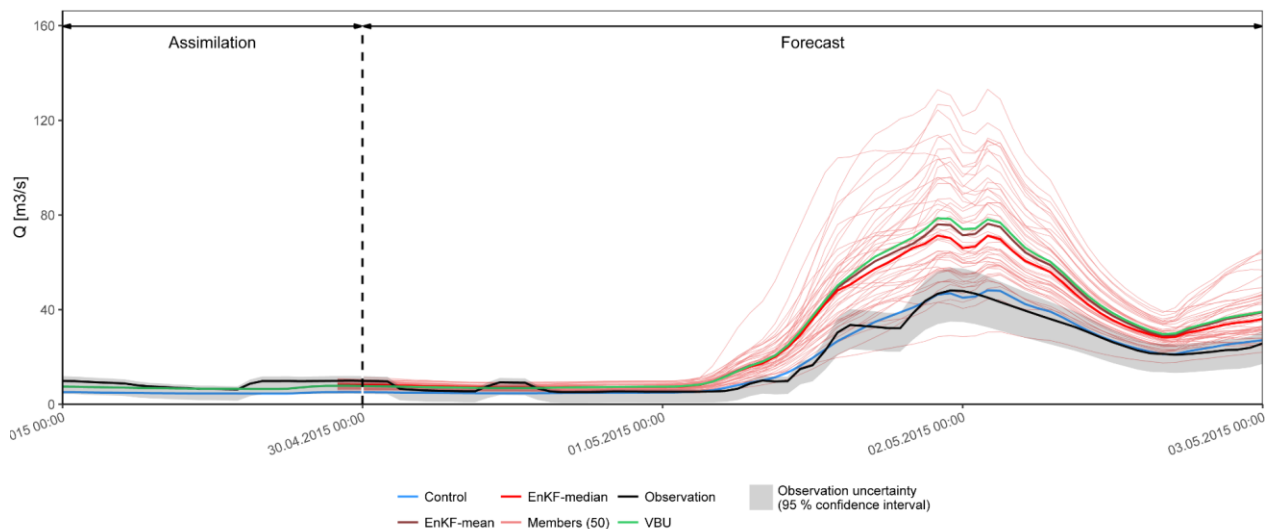


Figure 6: Forecasted discharges for the three different approaches at the Aigle station on the Grande-Eau River. The dashed line corresponds to 2015-04-30T00+01:00.

For the VBU approach, the external water discharge constantly causes an increase in the soil water content to try to reproduce the observed discharge. The longer the operation of the turbine over the last 24 hours, the more VBU will tend to increase the soil saturation. For the EnKF, the behavior slightly differs, mostly due to the consideration of the external input in the approach, by modifying the observed discharges with the turbine capacity for 50% of the members. When the observed discharge includes external turbined water, an overestimation is also observed in the forecasted discharge, even though it tends to be less extreme than the one of VBU (see Figure 6). On the contrary, when the turbine is not operating (i.e. the observed discharge corresponds to the natural discharge), the scheme tends to underestimate the observed discharge in the forecast, due to the discharge correction.

Results show that the implemented solution with a modification of the observed discharge for 50% of the members is not efficient. This issue needs further investigation to come to a robust system. Ideally, data from the turbine injecting the external water would be integrated directly in the model at the location of the hydropower plant to reduce this high source of uncertainty. Thereby, the performance would certainly be higher. In fact, tests have demonstrated that the EnKF approach can perform very well when applied over the hours preceding the studied event when the observed discharge corresponds to the natural discharge.

In addition, it must be noted that whereas the capacity of the turbine injecting the external water is fixed to $1.75 \text{ m}^3/\text{s}$, the observed discharge variations (increase and decrease) before the studied event vary between 3 and $4 \text{ m}^3/\text{s}$ for a base flow of about $6 \text{ m}^3/\text{s}$. Interestingly, an analysis of historical discharge data has revealed that for base flow of $2 \text{ m}^3/\text{s}$, the variations rather tend to be in the range of $1 \text{ m}^3/\text{s}$ for the same turbine operation. This will need to be better analyzed and will have to be considered for future development of the above presented EnKF methodology.

Furthermore, correct base flow estimates are important to achieve good results. If the model is not able to well reproduce the base flow, the correction of the EnKF might result in important over- or underestimation of the high flow events as a result of an inadequate correction of the soil water content. Therefore, implementing the presented methodology should only be considered on basins with a robust calibration.

5. CONCLUSIONS

Two data assimilation methods are explored in this study to improve the initial conditions of the model state variables with the objective of improving the quality of hydrological forecasts. They are compared with a reference scenario working without data assimilation (Control). The first method updates the soil saturation based on a volume-based update (VBU) over the 24 hours preceding the forecast. The second method is an implementation of the ensemble Kalman filter (EnKF). The semi-distributed conceptual hydrological model GSM-SOCONT is used for the simulations.

The methods are applied to two high flow events over two different catchments of the Upper Rhone River basin in Switzerland. Spatially interpolated precipitation data in combination with temperature data observed at stations are used both for the assimilation and forecast periods. The performance is evaluated with the Root Mean Square Error at hourly forecasting horizon up to 72 hours.

The results differ over the two basins. Over one basin, the two data assimilation methods provide better results than the Control, with EnKF outperforming the VBU approach. However, since the methodology is applied to a glacial basin where the glacial and snow melt reservoirs are not updated in the assimilation, the better performance might result from an overcompensation of the soil water content in the non-glacial part of the basin. Over the second basin, the results are completely different. The best performance is obtained with the Control simulation, thanks to a good reproduction of the observed discharge by the simulation without any correction. The low performance of the data assimilation approaches is explained by the impact of the water injection through a turbine from a reservoir external to the basin, considerably impacting the observed base flow and affecting the data assimilation. The effective turbine operation not being available to the system, the data assimilation techniques fail to adequately forecast the flow during the precipitation event.

ACKNOWLEDGEMENTS

The research project from which this article results is realized at the Ecole Polytechnique Fédérale de Lausanne (EPFL) in Switzerland, and is co-funded by the Centre de recherche sur l'environnement alpin (CREALP), located in Sion (Switzerland), and the Swiss Federal Office of Energy (SFOE). The authors express their gratitude to all involved stakeholders: the Canton of Valais, the Swiss Federal Office for Meteorology and Climatology (MeteoSwiss) and the engineering office HydroCosmos SA. The meteorological data used within the project come from the SwissMetNet network of MeteoSwiss, as well as the networks of the private company MeteoGroup Switzerland AG, the Intercantonal Measurement and Information System (IMIS) from the Institute for Snow and Avalanche Research (SLF), Agrometeo, the Canton of Bern, MeteoFrance, Electricite de France, the Regione Autonoma Valle d'Aosta and the ARPA Piemonte. The radar data have been provided by MeteoSwiss. We thank all the different data providers. Acknowledgments also go to the CREALP collaborators who cooperated on the MINERVE system and contributed to the set-up of the hydrometeorological databases used for the study. Thanks also to Charles de Broglie who contributed in the development of the data assimilation code used in this work.

REFERENCES

- Abaza, M., Anctil, F., Fortin, V. & Turcotte, R., 2015. Exploration of sequential streamflow assimilation in snow dominated watersheds. *Advances in Water Resources* 80, 79–89.
- Burgers, G., Jan van Leeuwen, P., Evensen, G., 1998. Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review* 126, 1719–1724.
- Clark, M.P., Rupp, D.E., Woods, R.A., Zheng, X., Ibbitt, R.P., Slater, A.G., Schmidt, J. & Uddstrom, M.J., 2008. Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in Water Resources* 31.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research* 99, 10143.

Foehn, A., García Hernández, J., Schaepli, B. & De Cesare, G., 2018. Spatial interpolation of precipitation from multiple rain gauge networks and weather radar data for operational applications in Alpine catchments. *Journal of Hydrology* 563, 1092–1110.

García Hernández, J. (2011). Flood management in a complex river basin with a real-time decision support system based on hydrological forecasts, *Communication 48 du Laboratoire de Constructions Hydrauliques*, Ed. A. Schleiss, EPFL, Lausanne.

García Hernández, J., Claude, A., Paredes Arquiola, J., Roquier, B. & Boillat, J.-L. (2014). Integrated flood forecasting and management system in a complex catchment area in the Alps – implementation of the MINERVE project in the Canton of Valais. In: Schleiss, A., Speerli, J., Pfammatter, R. (Eds.), *Swiss Competences in River Engineering and Restoration*. CRC Press, pp. 87–97 ISBN 978-1-138-02676-6 et 978-1-4987-0443-4.

García Hernández, J., Paredes Arquiola, J., Foehn, A. & Roquier, B. (2019). RS MINERVE – Technical manual v2.17. RS MINERVE Group, Switzerland.

Jordan, F. (2007). Modèle de prévision et de gestion des crues: optimisation des opérations des aménagements hydroélectriques à accumulation pour la réduction des débits de crue, *Communication 29 du Laboratoire de Constructions Hydrauliques*, Ed. A. Schleiss, EPFL, Lausanne.

Kalman, R.E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering* 82, 35.

R Core Team, 2018: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL <https://www.R-project.org>

Schaepli, B., Hingray, B., Niggli, M. & Musy, A., 2005. A conceptual glacio-hydrological model for high mountainous catchments. *Hydrology and Earth System Sciences* 9, 95–109.

Srikanthan, R., Amirthanathan, G. & Kuczera, G. (2008). Application of Ensemble Kalman Filter for flood forecasting in Australian rivers. *Australasian Journal of Water Resources* 12, 245–255.

Weerts, A.H. & El Serafy, G.Y.H. (2006). Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models: Particle and ensemble Kalman filtering. *Water Resources Research* 42.

Wickham, H., 2016: ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, URL <http://ggplot2.org>.

Wickham, H., 2018: htr: Tools for Working with URLs and HTTP. URL <https://CRAN.R-project.org/package=htr>, r package version 1.4.0.