

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

Alain, Foehn<sup>1</sup>

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

Damiano, Pasetto

Laboratoire d'écohydrologie, Faculté de l'Environnement Naturel, Architectural et Construit (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Suisse  
damiano.pasetto@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 factor for hydrological forecasts is the establishment of 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 station are assimilated. In the first approach, a volume-based update (VBU) compares the simulated and observed volumes over the past 24 hours to compute a correction factor used to update the 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 precipitation, temperature and discharge data. The comparison is carried out over 2 sub-basins of the Upper Rhône River basin upstream of Lake Geneva, where the MINERVE flood forecasting and management system is implemented. Results differ over the two studied basins. In one basin, the two data assimilation perform better than the Control simulation with the lowest error given by the VBU up to a forecast horizon of 35 hour and by the EnKF for higher forecast horizons. In the second basin, EnKF gives the lowest error over the few first hours of forecast, but then provides the weakest performance. The lowest error is given by the Control simulation, because the model already performs very well on the event without data assimilation.*

---

<sup>1</sup> Corresponding author

## 1. INTRODUCTION

Severe flooding events in recent decades have increased the need for reliable forecasting system 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 [Jordan2008, García2001, García2014].

Despite efforts on the calibration of the model, the 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 the uncertainties must be considered when analyzing the performances of a hydrological forecasting system.

Data assimilation techniques are mathematical tools developed to operationally correct the model results using the newly 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. Among them, the Ensemble Kalman filter (EnKF) is a Monte Carlo approach of the KF, performing an ensemble of model runs [Evensen1994]. Weerts et al. 2006 compared the EnKF with Particle filtering and showed that EnKF was more robust and outperformed the two other analyzed filters. In this study, an implementation of 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 simulation of the streamflow at the outlet of two sub-catchments of the Upper Rhone River basin over which an episode of high flow occurred over the period of availability (in 2012 for Reckingen and 2015 for Aigle). The characteristics of the sub-catchments are given in **Table 1**.

Station	River	Area [km <sup>2</sup> ]	Catchment mean elevation [m asl.]	Glaciation [%]	Return period of the studied event [years]
Reckingen	Rhone	214	2305	11.8	36
Aigle	Grande Eau	132	1562	0.8%	10

**Table 1:** Main characteristics of catchments

Both rivers are equipped each with a small run-of-river hydropower plant. The one over the Rhone has an equipped discharge of 5.7 m<sup>3</sup>/s (2012 yearly average: 10.6 m<sup>3</sup>/s). 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 m<sup>3</sup>. 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 external reservoir is turbined in a third more upstream hydropower plant with a capacity of 1.75 m<sup>3</sup>/s (2015 yearly average: 4.02 m<sup>3</sup>/s).

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 stations data into a 1km x 1km gridded data with a one hour temporal resolution.

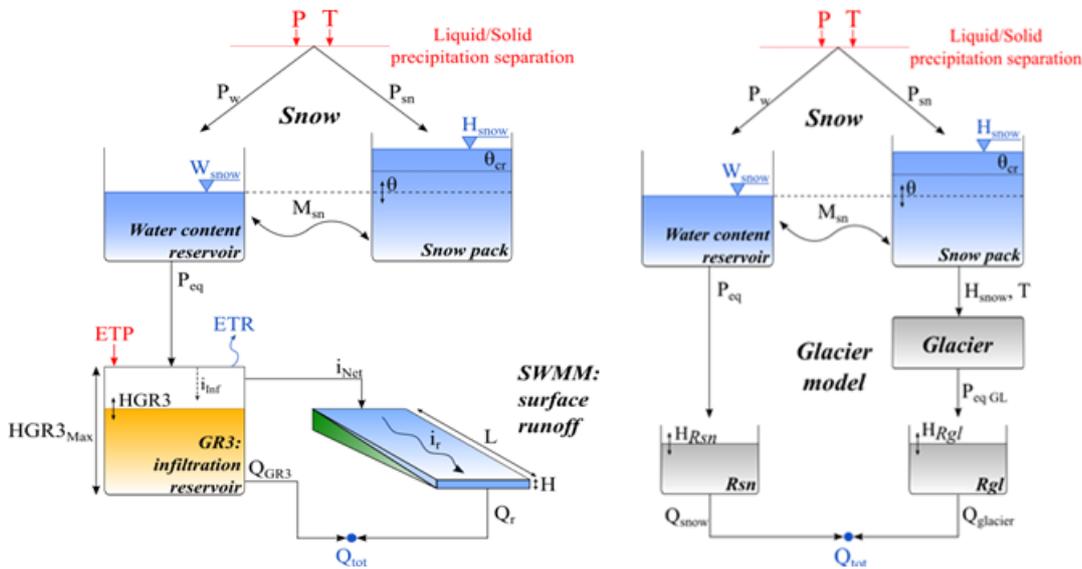
The temperature data are taken from a spatial product of temperature obtained through a kriging with external drift using elevation. The vertical lapse rate for temperature is computed each hour using a moving window of 150 m of altitude range and applying a signal filter. Most of the time, the temperature dispersion

at a given altitude is limited and the points are close to the regression line. However, this is not always true and the temperature distribution with altitude can be much more spread. In that case, the quality of the interpolated data is lower with a higher uncertainty.

In order to evaluate the performance of the data assimilation method, the mentioned spatial products of precipitation and temperature are used both for the assimilation and forecast simulation. Therefore, the simulations are based on observed data and are not subject to the uncertainty of meteorological forecasts. However, because uncertainties in these input data remain high in particular due to the complex topography of the Upper Rhone River basin, perturbed inputs are considered in the EnKF simulations (see Section 3.3).

## 2.2 GSM-SOCONT model

Simulations are performed with the rainfall-runoff semi-distributed conceptual GSM-SOCONT (Glacier and SnowMelt – Soil CONTRibution) model [Schaeffli2005, GarcíaHernández2019], illustrated in **Figure 1**. The main parameters of the model are given in **Table 2**.



**Figure 1:** Parameters of the SOCONT rainfall-runoff model ([GarcíaHernández2019])

Parameters	Units	Model SOCONT
Asn	mm/d/°C	Reference degree-day snowmelt coefficient
HGR3 <sub>Max</sub>	m	Maximum height of infiltration reservoir
KGR3	1/s	Release coefficient of infiltration reservoir
Kr	m <sup>1/3</sup> /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

**Table 2:** Main parameters of models SOCONT and GSM

The studied catchment is divided into elevation bands with an elevation range not exceeding 500 m. This allows in particular 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 melting 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.

## 2.3 Data assimilation techniques

Two data assimilation 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 3 years in the MINERVE system. The second method consists in an Ensemble Kalman filter, where an ensemble of possible model trajectories are corrected by assimilating the last 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 as measured in real-time. This step would improve the performance of the model during the forecast. To do so, the state variables of the model are updated using the available system observations.

In the GSM-SOCONT model, the state variables that 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. In the following we will indicate with  $\mathbf{x}_t$  the system state vector of dimension  $nstate=3 \times nnodes$ , which elements are the state variables HGR3, Hr, 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),  $\mathbf{w}_t$  is the possible occurrence of system noise (not directly considered in this analysis).

The assimilation procedure is performed using real measurements of the system, discharge data based on water level observations at the outlet of the watershed, provided by the Swiss Federal Office for the Environment (FOEN). To do so, the initial discharge of the (kinematic wave) river reaches (Qini) of the model are also considered as state variables. 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). 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 we describe the three assimilation procedures considered here to produce the forecasts of discharge.

### 2.3.1 Control simulation (Control)

The reference scenario is computed by running the model (eq. 1) without any perturbation of the input data (i.e.,  $\mathbf{u}_t$  corresponds to the nominal values of precipitation and temperature) and without considering the discharge measurements performed 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 updated state variable is the soil moisture (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 flows 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 results. Note that VBU does not take into consideration the uncertainties in the input, model, and observations. The update takes into consideration all the observations collected during the 24 h 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 Ensemble Kalman filter (EnKF), which better allows the update of the state variables taking into consideration model and measurements uncertainties. 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. EnKF is based on the sequential repetition of two steps: the prediction (or forecast) step and the analysis (or update) step. In the prediction step the different members of the ensemble are independently advanced in time by running the model (eq. 1) using different random samples of the uncertain forcing terms:[Srikanthan2008, Abaza2015]:

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

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

In the analysis step, the predicted state variables  $\mathbf{x}_{t+1}^{i,f}$  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:

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

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

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

and it is used to compute the model error for each ensemble member  $i$ :

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

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

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

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:

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

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

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

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 consisted 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.

## 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 ensures spread in the ensemble. Ensembles are initialized with a perturbation of the main state variables of the model.

### 3.1.1 Precipitation uncertainty

Precipitation measurements are a large source of uncertainty, here described with a temporally-correlated lognormal random variable. During the prediction step of EnKF, the nominal value of precipitation is perturbed with a multiplier coefficient as follows:

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

with

$$e_t^i = \exp(z_t^i) \sim \log N(1, \sigma_e) \quad (10b)$$

where  $P_t$  is the measured precipitation at time  $t$ ,  $e_t^i$  is the multiplier coefficient for the  $i^{\text{th}}$  member at time  $t$ ,  $P_t^i$  is the perturbed precipitation for the  $i^{\text{th}}$  member at time  $t$  and  $z_t^i \sim N(m_z, \sigma_z^2)$ .

At time  $t=0$ ,  $z_0^i$  is sampled from  $z_0^i \sim N(m_z, \sigma_z^2)$ . Parameters  $m_z$  and  $\sigma_z$  are computed imposing that  $m_e = 1$  mm/h and  $\sigma_e = 0.7$ . This value of  $\sigma_e$  is in agreement with the Root Mean Squared Error (RMSE) computed in a previous validation study [Foehn2018] based on the cross-validation of a spatial product of precipitation over 4 events of heavy precipitation over the basin.

At time  $t>0$ , in order to ensure a temporal correlation of the perturbation in a given member, the time evolution of model errors is simulated as follows [Clark2008]:

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

$$\omega_t^i \sim N(0, \sigma_z^2) \quad (11b)$$

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

where  $\omega_t^i$  is the sample white noise,  $\rho$  is the temporal persistence parameter,  $\Delta T$  is the simulation time step and  $\tau$  the decorrelation time step. In the present study,  $\Delta T$  is one hour and  $\tau$  is fixed to 24 hours (following Clark2008).

### 3.1.2 Temperature uncertainty

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

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

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

For time  $t>0$ , a temporal correlation of the perturbation is considered following Eq. (9a) to (9c). The decorrelation time is fixed to 12 hours for the temperature.

### 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  $\varepsilon_y$  is a hyper-parameter allowing to define the perturbation proportionally to the discharge values and  $\gamma$  corresponds to a correction for possible external discharge contributions to the basin, like in the case of the Grande-Eau (1.75 m<sup>3</sup>/s from the Lac d'Arnon located outside of the basin). The  $\gamma$  correction is randomly sampled to represent the installed discharge capacity of the diversion for half of the members and to 0 m<sup>3</sup>/s for the other members.

### 3.1.4 Members initialization

Before the first assimilation, initial conditions of the state variables are also perturbed. Level in infiltration reservoir, level of surface runoff and snow water equivalent height 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.

### 3.2 Performance evaluation

Here we consider two precipitation events and we use VBU and EnKF to assimilate streamflow measurements every 2 hours. After each assimilation, a new streamflow forecast of 3 days is produced using the actual precipitation and temperature, so that the results are not subject to the uncertainty due to meteorological forecasts. The performances of the control scenario, VBU and EnKF are evaluated by computing the root mean squared error (RMSE) between the forecasted and measured streamflows at different lead times  $L$  (each hour) during the forecast as follow:

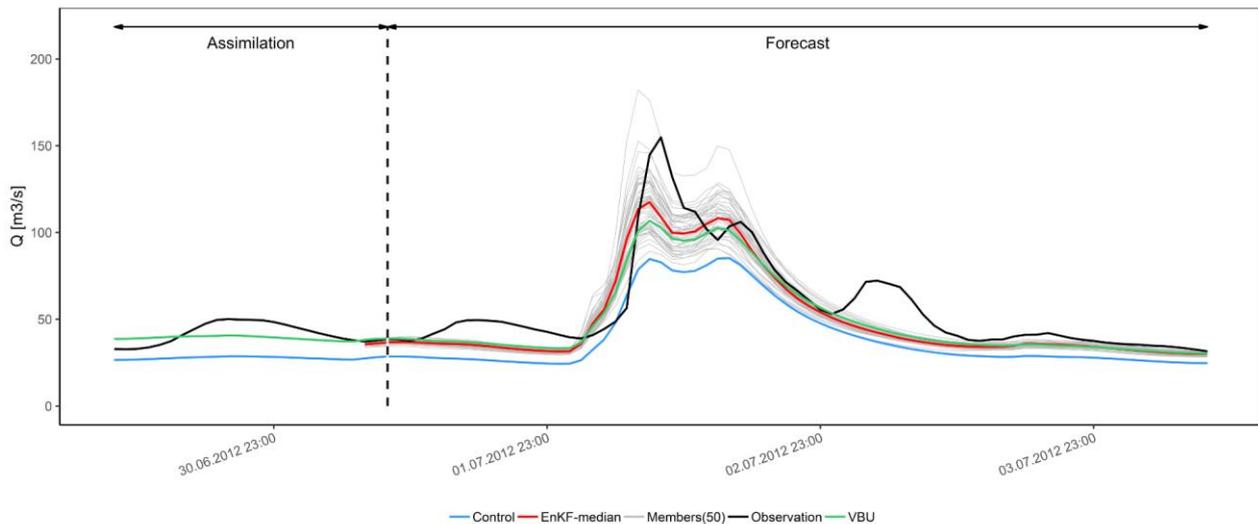
$$RMSE_L = \sqrt{\frac{\sum_{j=1}^n \{Q_{L,j}^f - Q_{L,j}^{obs}\}^2}{n}} \quad (12)$$

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

## 4. RESULTS AND DISCUSSION

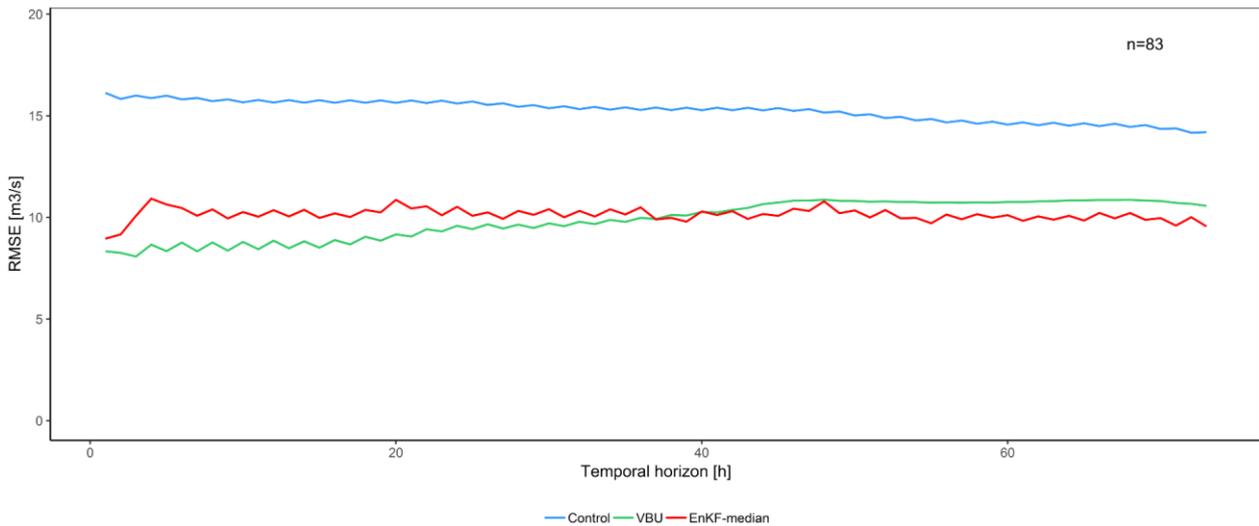
Simulations have been run for the two events presented in **Table 1**.

**Figure 2** shows the forecasted flow about 24 hours before the peak flow for the high flow event at station Reckingen (Rhone River). Over the assimilation period, the difference between the VBU (green line) and the Control (blue line) are well visible, with the VBU line much closer to the observations. The red line, existing only over the two hours preceding the assimilation time (corresponding to the propagation of the members) is close to the observed discharge value of the assimilation time (dashed line).



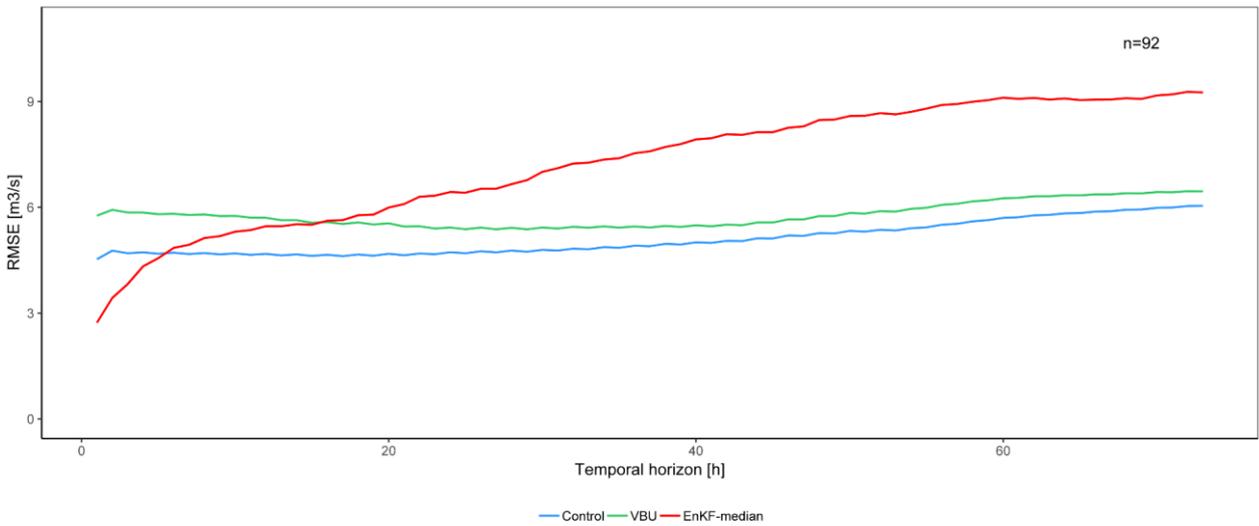
**Figure 2:** Forecasted discharges for the three different approaches at the Reckingen station on the Rhone River. The dashed line corresponds to 2012-07-01T10+01:00.

The RMSE values obtained for the event are presented in **Figure 3**. The two assimilation methods clearly outperform the Control simulation for the entire forecast period (72 hours). Interestingly, the VBU approach provides better performances up to a forecast horizon of 40 hours. For higher temporal horizons, the EnKF approach gives the lowest error.



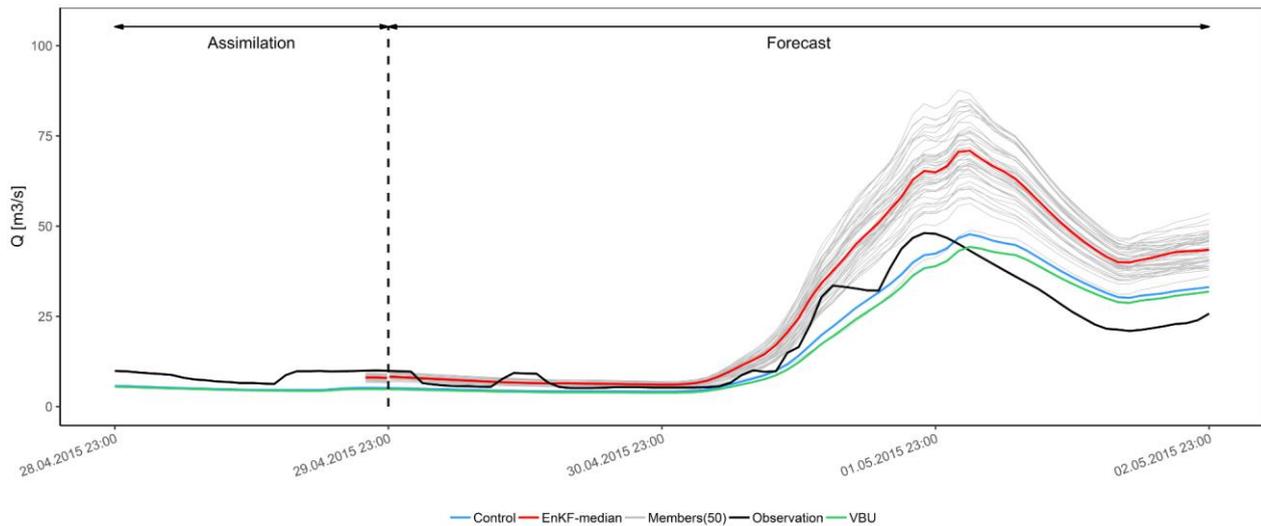
**Figure 3:** RMSE values obtained for the three approaches for the high flow event at the Reckingen station on the Rhone River. The n value indicated the number of run considered in the computation of the RMSE value of each temporal horizon.

Over the second studied high flow event, concerning the Grande-Eau basin, results are quite different (Figure 4). In that case, the Control simulation provides the best performance except for the few first hours, where EnKF performs best. Rapidly, the error of the EnKF increases with temporal horizon and exceeds the error of the VBU approach.



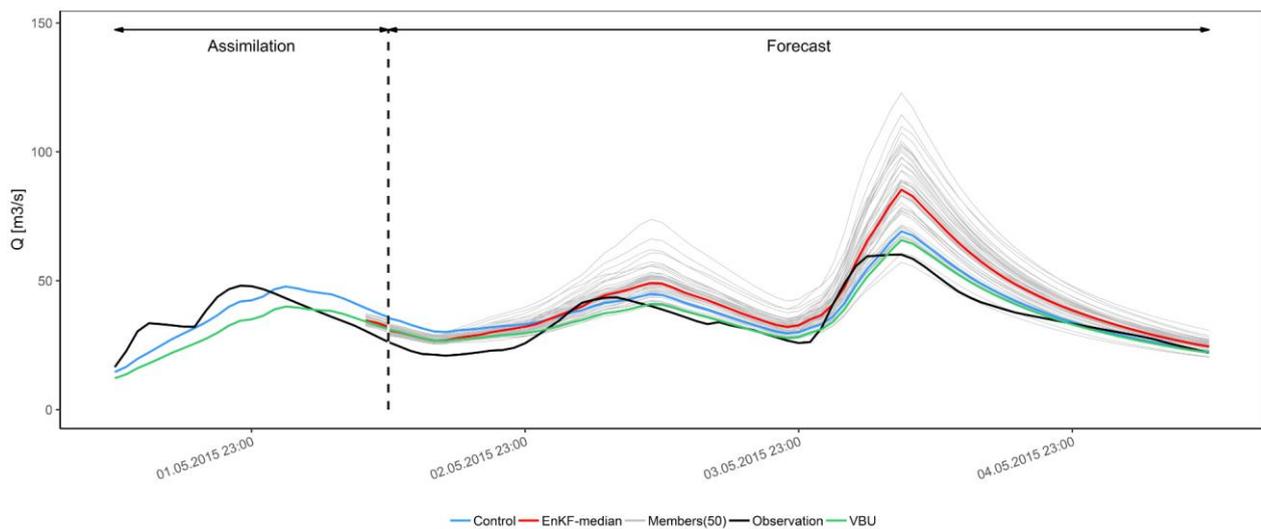
**Figure 4:** RMSE values obtained for the three approaches for the high flow event at the Aigle station on the Grande-Eau River. The n value indicated the number of run considered in the computation of the RMSE value of each temporal horizon.

To better understand the reason of this low performance of the EnKF, **Figure 5** illustrates the discharge simulated with the forecast generated about 36 hours before the start of the precipitation. The discharge forecasted by the EnKF approach is closer to the observed discharge over the first 30 hours of the forecast, the Control and VBU underestimating the base flow. Indeed, the EnKF has increased the soil saturation to have a base flow corresponding to the discharge given by the data. When the precipitation starts, the discharge forecasted by the EnKF overestimates the peak flow, in addition to a temporal delay of the simulated peak for all three approaches. An analysis of the initial conditions of the Control and the VBU has revealed that the VBU does not approach more the observed discharge values because it reaches 75% of soil saturation over the entire basin (constraint fixed in the VBU approach).



**Figure 5:** 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.

The overestimation by the EnKF approach remains valid over the entire event, composed of three successive fronts of precipitation that resulted in three successive increase of the discharge (**Figure 6**).



**Figure 6:** Forecasted discharges for the three different approaches at the Aigle station on the Grande-Eau River. The dashed line corresponds to 2015-05-02T12+01:00.

Further investigation will be required to understand the reason of this overestimation. However, one can already say that likely, the inadequacy of the model to reproduce the low flow before the precipitation event results in a correction by the EnKF that does not correspond to the state of the soil before the event. The correction brought by the EnKF probably results in an excessive increase of the soil saturation. This excessive increase might be linked to the low maximum height of the infiltration reservoir which is only of 0.2 m. Therefore, a slight change in the soil saturation can have a considerable effect on the basin reaction when receiving precipitation.

## 5. CONCLUSION

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. The first method updates the soil saturation based on a volume-based update (VBU) over the 24 hours preceding the forecast. The second 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 events of high flow events over two different catchments of the Upper Rhone River basin in Switzerland. Spatial products of observed precipitation and temperature data are used both for the assimilation and forecast parts. 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 result than the Control simulation corresponding to a simulation of the model without data assimilation. The VBU performs better up to a forecast horizon of 35 hours, above which the EnKF then results in lower errors. Over the second basin, the results are completely different. The EnKF performs best over the first few hours but then provides the highest error. The best performance is obtained by the Control simulation, thanks to the ability of the model to reproduce the event without any assimilation. The EnKF results in a clear overestimation of the discharge, resulting from an adaption of the soil saturation to reproduce the base flow that affects the ability of the model to absorb part of the incoming precipitation.

## 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), the Universitat Politècnica de Valencia 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 AND CITATIONS

- 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.
- 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.
- 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.

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

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.