International Conference on Hydroinformatics

2014

Moving Horizon Estimation To Assimilate Snow And Soil Moisture Data Into The HBV Hydrological Model

Rodolfo Alvarado Montero

Dirk Schwanenberg

Peter Krahe

Aynur Şensoy

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

MOVING HORIZON ESTIMATION TO ASSIMILATE SNOW AND SOIL MOISTURE DATA INTO THE HBV HYDROLOGICAL MODEL

RODOLFO ALVARADO MONTERO (1), DIRK SCHWANENBERG (1,2), PETER KRAHE (3), AYNUR ŞENSOY (4)
(1): Institute of Hydraulic Engineering and Water Resources Management, University of Duisburg-Essen, Universitätsstrasse 15, Essen, 45141, Germany
(2): Deltares, Rotterdamseweg 185, Delft, 2629HD, The Netherlands
(3): Federal Institute of Hydrology, Department of Water Balance, Forecasting and Predictions, Koblenz, 56068, Germany
(4): Department of Civil Engineering, Faculty of Engineering, Anadolu University, Eskisehir, 26555, Turkey

We present a variational data assimilation approach based on a Moving Horizon Estimation (MHE) applied to the HBV hydrological model. This framework enables the modification of the model inputs precipitation and temperature as well as the model states soil moisture, upper zone storage and lower zone storage. It considers data products for snow cover, snow water equivalent and soil moisture and observed streamflow.

The performance of the framework is evaluated for three test sites: i) the data–dense catchment of the upper Main River (2419 km^2), Germany, for which the HBV model already produces excellent results, ii) a comparable upstream catchment of the Nahe River (1468 km^2), Germany, and iii) a data-sparse environment in the upper basin of Karasu River in Turkey (10,275 km^2). The added value of the data assimilation approach is relatively limited in the case of (i) and (ii), but more substantial for the data-sparse environment (iii) with only a limited amount of operational ground data.

INTRODUCTION

Data assimilation plays an important role in real-time flow forecasting systems. The main purpose is to provide updated model states using recent observations. These states are then used as initial conditions for the subsequent forecasts to achieve a better forecast lead time accuracy. This concept is referred to as sequential data assimilation [14]. The basic idea behind it is to combine observations and model simulations in an optimization problem that improves the quality of both observed and simulated data [11]. There are various data assimilation techniques, of which the most commonly used are the Ensemble Kalman Filter (EKF), the Particle Filter and variational methods [1][3][4][5][10]. The last approach is essentially an optimization procedure that adjusts uncertain variables and/or parameters to obtain the best fit between model states and observations.

Applications of spatially distributed data assimilation techniques in hydrological models have been actively studied using different types of approaches [7][9][11]. To provide the hydrologic community with relevant products derived from the raw satellite observations, the *European Organization for the Exploitation of Meteorological Satellites* (EUMETSAT) established the *Satellite Application Facility on Support to Operational Hydrology and Water Management* (H-SAF) in 2005. The objectives of H-SAF are twofold: i) provide products derived from existing and planned satellites for operational hydrology in adequate space-time resolution, ii) validate the products by means of comparing the satellite-derived observations with radar measurements and with data from synoptic observation networks, by assimilating the satellite-derived products into hydrological models and by assessing the impacts of the products on hydrological applications. The currently available and operational H-SAF products include information about precipitation, snow and soil moisture conditions.

Data assimilation is used in this study to integrate remote sensing data into hydrological models to improve the lead time performance of streamflow forecasts in the context of operational hydrological forecasting systems. The purpose here is to present a variational approach based on a Moving Horizon Estimation (MHE) applied to the HBV hydrological model. The novel framework enables the modification of the HBV inputs precipitation and temperature as well as the model states soil moisture and the upper and lower storage terms of the conceptual model. It aims at the future integration of H-SAF data products for snow cover, snow water equivalent and soil moisture. The main advantage of this approach is the highly flexible formulation of distance metrics for the introduction of noise into the model and the agreement between simulated and observed variables as well as its robustness regarding non-equidistant, noisy and partially missing data.

METHODOLOGY

Our dedicated implementation of the HBV model is documented in Schwanenberg et al. [12] and follows the methodology of Bergström [2] and Lindstrom et al.[8]. It strictly considers an implementation according to:

$$x^k = f(x^{k-1}, d^k, u^k) \tag{1}$$

$$y^k = g(x^k, d^k, v^k) \tag{2}$$

where *x*, *y*, *d* are the state, output and external forcing vectors, respectively, *u*, *v* are noise terms, *f*, *g* are functions representing arbitrary linear or nonlinear components of the HBV model and *k* is the time step index.

Based on Eq. (1)-(2) above, we formulate the Moving Horizon Estimation (MHE) for a forecast time $k = 0$ over an assimilation period $k = [-N+1, 0]$ of $N \geq 1$ time steps by an optimization problem according to:

$$\min_{u,v} \sum_{k=-N+1}^{0} \left[w_x \left\| \hat{x}^k - x^k(u) \right\| + w_y \left\| \hat{y}^k - y^k(u,v) \right\| + w_u \left\| u^k \right\| + w_v \left\| v^k \right\| \right] \tag{3}$$

$$\text{subject to} \quad u_L \leq u^k \leq u_U$$

$$\mathrm{v}_L \leq v^k \leq v_U \tag{4}$$

where $\hat{x}^k, \hat{y}^k$ are observations of the state and the dependent variable vectors, $\|.\|$ is a suitable norm penalizing the deviation between observed and simulated quantities and the introduction of noise by the data assimilation procedure, $w_{x,y,u,v}$ are weighting coefficients for defining the trade-off between different penalties. Furthermore, the noise terms get bounded by inequality constraints. For the sake of simplicity, our formulation considers constant bounds only.

The key to the efficient solution of the optimization problems above, in particular in operational applications with runtime restrictions, is the computation of the derivatives of the objective function we refer to as $J(u, v)$ for applying gradient-based optimizers such as IPOPT [13]. Since numerical differentiation is a computational burden for larger optimization problems and introduces truncation errors, we rely on adjoint modelling based on algorithmic differentiation in reverse mode for tracing back first-order derivatives backwards in time through the model [6].

To evaluate the data assimilation procedure, we distinguish the assessment of model performance and forecast lead time performance. The first one is used for evaluating the model calibration and validation. Furthermore, it serves as a measure of the potential impact of a specific assimilation on model outputs under consideration of the model structure. For example, the modification of a zero precipitation does not contribute to a state update if the simulated streamflow is already too high. The second one shows benefits of data assimilation procedures and their updated model states in the forecasts results. We use:

$$BIAS = \frac{1}{N}\sum_{k=-N+1}^{0} \hat{x}^k - x^k \tag{5}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=-N+1}^{0} \left(\hat{x}^k - x^k\right)^2} \tag{6}$$

$$R2 = \left[\sqrt{\sum_{k=-N+1}^{0} \left(x^k - \bar{x}\right)^2}\sqrt{\sum_{k=-N+1}^{0} \left(y^k - \bar{x}\right)^2}\right]^{-1} \cdot \sum_{k=-N+1}^{0} \left(x^k - \bar{x}\right)\left(y^k - \bar{y}\right) \tag{7}$$

$$NSE = 1 - \left[\sum_{k=-N+1}^{0} \left(y^k - \bar{y}\right)^2\right]^{-1} \cdot \sum_{k=-N+1}^{0} \left(y^k - x^k\right)^2 \tag{8}$$

where BIAS is the bias between observation $\hat{x}^k$ and simulation x^k, RMSE is the Root Mean Square Error, R2 is the correlation coefficient and NSE is the Nash-Sutcliffe model efficiency. For assessing the forecast lead time accuracy, we reformulate the indicators in Eq. (5)-(8) according to the example:

$$BIAS^L = \frac{1}{N}\sum_{k=-N+1}^{0} \hat{x}^k - x^{k,L} \tag{9}$$

where L is the forecast lead time we want to assess and the value $x^{k,L}$ indicates the value of a forecast with a forecast time of k-L.

RESULTS

Study Areas and Model Setups

The data assimilation method is tested for three catchments. We select two headwater catchments in Germany as representatives of a data-rich environment: The first is a headwater catchment of the River Main (Main1), located upstream of gauge Schwürbitz with a drainage area of about 2419 km², of which 40% is covered with forest. Elevation ranges between 250 m and 1100 m above sea level (ASL). The mean annual discharge at gauge Schwuerbitz is 30.1 m³/s for the observation period 1941-2009. The second is the headwater catchment of the River Nahe (Nahe1), which is located upstream of gauge Martinstein. It has a drainage area of 1468 km² with 60% covered by forest and an elevation between 150 m and 800 m ASL. The mean annual discharge at gauge Martinstein for the period 1963-2011 is 15.8 m³/s. Both catchments are located in low mountain ranges with no considerable groundwater supply. Our third case is the Karasu catchment in Turkey, located upstream of gauge Kemah, in a data-sparse environment, with an area of 10,275 km^2. Main land types are pasture, shrub, grass and wasteland. Elevation ranges in altitude from 1125 to 3487 m ASL.

Calibration and Validation Results

The HBV model performance in the calibration and validation period is shown in Table 1. The two catchments located in Germany have a better model performance than the Turkish case. We suspect this is because of the much better data availability both in time and space scales.

Table 1. Model performance in calibration and validation periods (1962-2006 / 2007-2012 for the German test sites, 2001-2008 / 2009-2012 for the Turkish basin)

Basin	Av. Flow	Calibration				Validation			
	Q	BIAS	RMSE	R2	NSE	BIAS	RMSE	R2	NSE
	[m^3/s]	[m^3/s]	[m^3/s]	[-]	[-]	[m^3/s]	[m^3/s]	[-]	[-]
Karasu	85.14	-1.49	33.22	0.840	0.840	-6.69	34.07	0.75	0.74
Main1	31.05	1.37	11.26	0.912	0.909	-1.22	14.21	0.85	0.85
Nahe1	15.65	-0.43	6.858	0.917	0.917	-1.72	8.14	0.87	0.87

Assimilation Results

The first experiment tests the response of the model to the assimilation of individual variables, such as precipitation, by assessing the maximum agreement between observed and simulated streamflow. Therefore, we permit a large variation of the variable and place a high emphasis on streamflow deviations. The experiment is also used as a technical verification of the proper functioning of the data assimilation procedure.

Table 3 shows the impact of all the assimilation variables in each basin. The highest impact on the assimilation procedure is achieved by using the upper zone state and a combination of all variables. The biggest improvement is found in the Karasu catchment, in which NSE shifts from 0.839 without data assimilation to 0.987 using only assimilation of the upper zone state. Main1 and Nahe1 catchments already perform above 0.90 without data assimilation, although the impact of assimilation increases to a maximum to an almost perfect fit between observed and simulated streamflow.

It is obvious that model modifications in components which are closer to the model's response are more effective in terms of the model performance improvement. Since this is not a value in itself, the next experiment will assess if the modifications lead to a better lead time performance of the forecasts.

Table 2. Model performance of a simulation run without data assimilation in comparison to a run with different data assimilation setups

Basin	**Mean flow [m³/s]**	**Perf. Ind.**	**Without DA**	**DA (ΔP)**	**DA (ΔT)**	**DA (ΔSM)**	**DA (ΔUZ)**	**DA (ΔLZ)**	**DA (ALL)**
Karasu	84.99	BIAS	-1.49	-1.51	-2.82	-0.10	0.77	1.34	-0.06
		RMSE	33.22	19.05	15.61	16.33	9.38	21.32	3.58
		R2	0.843	0.948	0.966	0.961	0.987	0.934	0.998
		NSE	0.839	0.947	0.965	0.961	0.987	0.934	0.998
Main1	31.05	BIAS	1.372	0.369	1.227	-0.853	0.401	0.2	0.038
		RMSE	11.261	6.358	7.177	8.393	4.425	5.813	1.729
		R2	0.912	0.971	0.964	0.951	0.986	0.976	0.998
		NSE	0.909	0.971	0.963	0.950	0.986	0.976	0.998
Nahe1	15.65	BIAS	-0.431	-0.183	-0.36	-0.815	0.077	0.11	-0.008
		RMSE	6.858	3.467	4.905	5.117	1.735	3.395	1.093
		R2	0.917	0.979	0.958	0.956	0.995	0.980	0.998
		NSE	0.917	0.979	0.958	0.954	0.995	0.980	0.998

In a next step, we conduct a hindcast experiment for the Main1 model using a period of three years (Dec 2003-Dec 2006) and assess the lead time performance of forecasts based on assimilated system states generated by the procedure above.

The lead time dependent Mean Average Error (MAE) is presented in Figure 1. The constant value of approximately 7.1 m^3/s shows the performance without data assimilation and represents the reference for the different assimilation setups. All runs show an improvement of the MAE over all lead times with respect to the case without assimilation, except for the one which assimilates soil moisture. Best results are achieved by the assimilation of the upper and lower zone states.

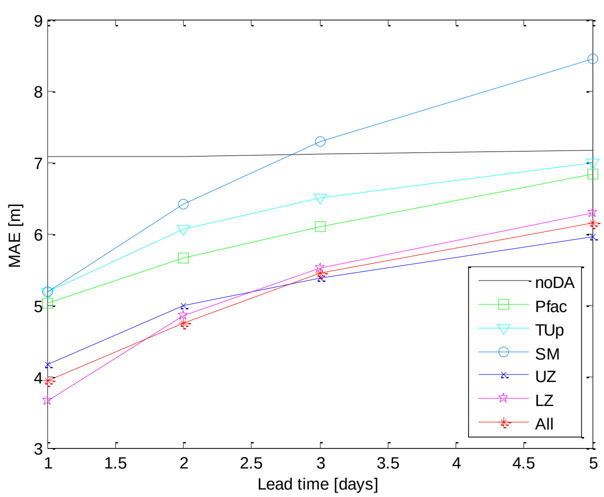

Figure 1. Mean Absolute Error (MAE) at different lead times for individual assimilation variables of the Main1 hindcast in the period Dec 2003 – Dec 2006.

Potential Benefits using H-SAF Products

The use of H-SAF products allows the integration of observations of snow coverage, snow water equivalent and soil moisture in the assimilation procedures. These terms are introduced in the objective function by defining a norm for the agreement between observations and model simulations. Since the time overlap between validated observations and H-SAF products is still small, we assess the potential benefit of these products by another experiment. First, we generate a set of "perfect" observations for snow coverage, snow water equivalent and soil moisture by using the existing model outputs. Then, we introduce noise to the model inputs. Finally, we conduct a hindcasting experiment and assess the lead time performance of the noise model in combination with the different data assimilation setups.

Figure 2 shows the largest failure of the assimilation procedure for a snow melt event in March 2006. Since the temperature is below zero in most elevation zones before March 23, the modification of precipitation does not lead to a significant improvement of the assimilated streamflow. However, the assimilation procedure increases precipitation significantly, to achieve at least small streamflow improvements. This results in a large increase of the snow water equivalent. The temperature suddenly increases on March 24 and triggers a considerably overestimated streamflow in the forecast due to excessive snow melt. Snow water equivalent and streamflow get back into a realistic range as soon as the assimilation procedure captures the observed discharge increase in the newer forecasts for March 25-26.

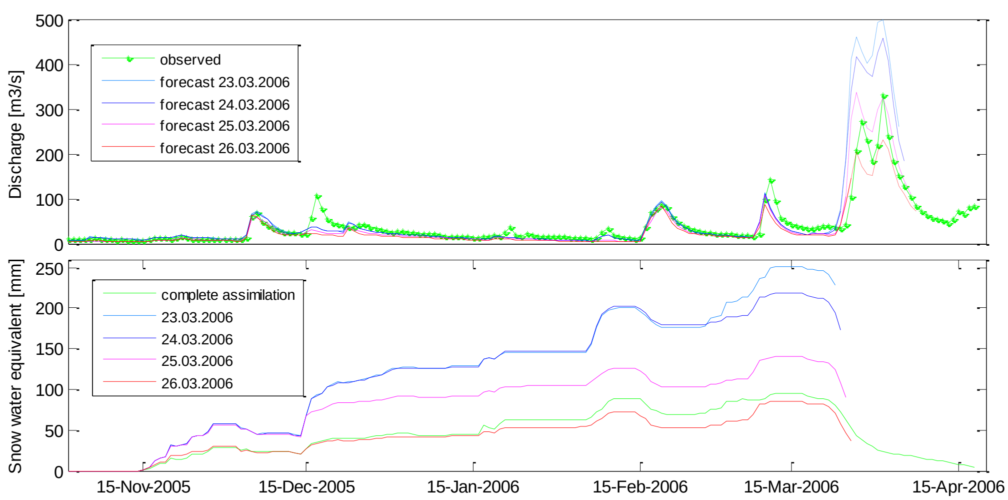

Figure 2. Lead time performance of the Main1 model for a flood event in March 2006 using observed discharge in the assimilation: i) above: comparison of observed and (assimilated) forecasted streamflow of forecast times between March 23-26, ii) below: comparison of "perfect" and assimilated snow water equivalent for the same forecasts.

Let us assume the availability of a basin-averaged, observed snow water equivalent. In this case, we can use the product to support the assimilation of precipitation. Because of the low sensitivity of the streamflow to a precipitation change, the assimilation procedure has much freedom to approximate the simulated snow water equivalent to the observation (Figure 3). The subsequent streamflow forecasts consider the adjusted states and show major improvements in the forecast lead time performance.

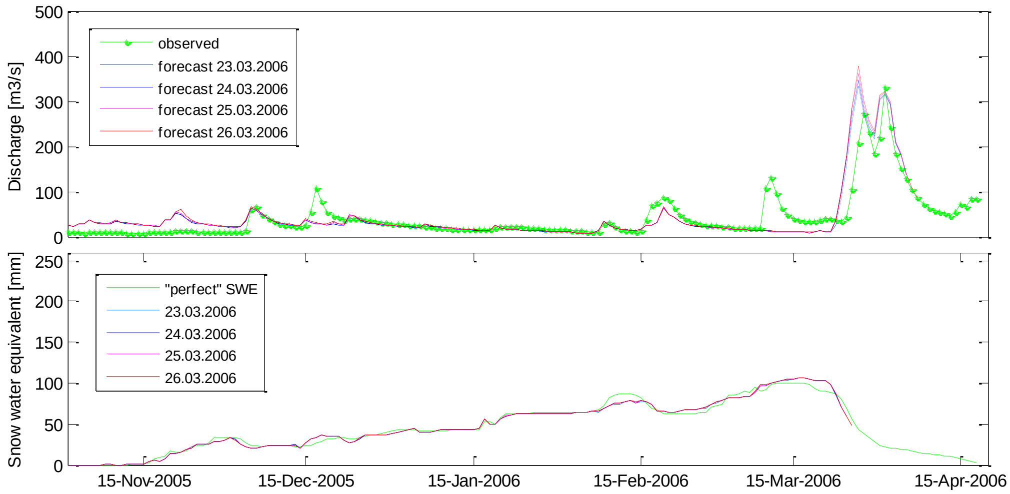

Figure 3. Lead time performance of the Main1 model for a flood event in March 2006 using observed snow water equivalent and discharge in the assimilation: i) above: comparison of observed and (assimilated) forecasted streamflow of forecast times between March 23-26, ii) below: comparison of “perfect” and assimilated snow water equivalent for the same forecasts.

CONCLUSIONS

A reimplementation of the HBV model with an adjoint mode for computing first-order derivatives enables its efficient application in variational data assimilation approaches such as the Moving Horizon Estimation. The combination of HBV and MHE forms a flexible framework to assimilate H-SAF products such as snow coverage, snow water equivalent and soil moisture into the model by modifying the model inputs precipitation and temperature as well as the soil moisture state. The additional availability of observed discharge enables a simultaneous assimilation of upper and lower zone states.

The novel framework has been successfully validated in several experiments and shows the potential benefit of the new H-SAF data products. Future research will focus on the further practical validation of these products in additional hindcasting experiments. Whereas the lead time improvements in data-dense environments such as the Main1 and Nahe1 catchments are limited, a significantly higher benefit is expected for data-sparse environments such as the mountainous Karasu basin.

REFERENCES

[1] Andreadis K.M., Lettenmaier D.P. “Assimilating remotely sensed snow observations into a macroscale hydrology model”, *Advances in Water Resources*, Vol. 29, (2006), pp 872-886.

[2] Bergström S. “Computer Models of Watershed Hydrology: The HBV model”. In: Singh, V.P. (Ed.), Water Resources Publications, Highlands Ranch, CO., pp. 443-476, (1995)

[3] Clark M.P., Rupp D.E., Woods R.A., Zheng X., Ibbitt R.P., Slater A.G., Schmidt J., Uddstrom M.J. "Hydological data assimilation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model", *Advances in Water Resources*, Vol. 31, (2008), pp 1309-1324.

[4] Dechant C., Moradkhani H. "Radiance data assimilation for operational snow and streamflow forecasting", *Advances in Water Resources*, Vol. 34, (2011), pp 351-364.

[5] Dumedah G., Coulibaly P. "Evaluating forecasting performance for data assimilation methods: The ensemble Kalman filter, the particle filter, and the evolutionary-based assimilation", *Advances in Water Resources,* Vol. 60, (2013), pp 47-63.

[6] Griewank A., Walther A., Evaluating Derivatives, second edition. SIAM, 2008.

[7] Han E., Merwade V., Heathman G.C. "Implementation of surface soil moisture data assimilation with watershed scale distributed hydrological model", *Journal of Hydrology*, Vol. 416-417, (2012), pp 98-117

[8] Lindstrom G, Johansson B, Persson M, Gardelin M, Bergstrom S, "Development and test of the distributed HBV-96 hydrological model", *Journal of Hydrology,* Vol. 201, (1997), pp 272–288.

[9] Liu, Y., Weerts, A.H., Clark, M., Hendricks Franssen, H.J., Kumar, S., Moradkhani, H., Seo, D.J., Schwanenberg, D., Smith, P., van Dijk, A.I J M., van Velzen, N., He, M., Lee, H., Noh, S.J., Rakovec, O., Restrepo, P. "Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities", *Hydrology and Earth System Science*s, Vol. 16, (2012), 3863–3887.

[10] McMillan H.K.., Hreinsson, E.Ö., Clark, M.P., Singh, S.K., Zammit, C., Uddstrom, M. J. "Operational hydrological data assimilation with the recursive ensemble Kalman filter" *Hydrology and Earth System Sciences*, Vol. 17, (2013), pp 21–38.

[11] Reichle R. "Data assimilation methods in the Earth sciences", *Advances in Water Resources*, Vol. 31, (2008), pp 1411-1418.

[12] Schwanenberg D, Sheret I., Rauschenbach T., Galelli S., Viera J.M., Pinho J.L. "Adjoint Modeling Framework for Water Resources Models", 10th International Conference on Hydroinformatics HIC 2012, Hamburg, Germany, 2012

[13] Wächter A. and Biegler L.T. "On the Implementation of a Primal-Dual Interior Point Filter Line Search Algorithm for Large-Scale Nonlinear Programming", *Mathematical Programming*, Vol. 106, (2006), pp 25-57.

[14] Weerts A., El Serafy G.Y., Hummel S., Dhondia J., Gerritsen H. "Application of generic data assimilation tolos (DATools) for flood forecasting purposes", *Computer and Geosciences*, Vol. 36, No.4, (2010), pp 453-463.