20th International Congress on Modelling and Simulation, Adelaide, Australia, 1–6 December 2013
www.mssanz.org.au/modsim2013

Impact of observation error structure on satellite soil moisture assimilation into a rainfall-runoff model

C. Alvarez-Garreton [a], **D. Ryu** [a], **A. W. Western** [a], **W. Crow** [b], **D. Robertson** [c]

[a] *Department of Infrastructure Engineering, The University of Melbourne, Parkville, Victoria, Australia*
[b] *USDA-ARS Hydrology and Remote Sensing Laboratory, Beltsville, Maryland, United States*
[c] *CSIRO Land and Water, Victoria, Australia*
Email: calverez@student.unimelb.edu.au

Abstract: In the Ensemble Kalman Filter (EnKF) - based data assimilation, the background prediction of a model is updated using observations and relative weights based on the model prediction and observation uncertainties. In practice, both model and observation uncertainties are difficult to quantify thus have been often assumed to be spatially and temporally independent Gaussian random variables. Nevertheless, it has been shown that incorrect assumptions regarding the structure of these errors can degrade the performance of the stochastic data assimilation.

This work investigates the autocorrelation structure of the microwave satellite soil moisture retrievals and explores how assumed observation error structure affects streamflow prediction skill when assimilating these observations into a rainfall-runoff model. An AMSR-E soil moisture product and the Probability Distribution Model (PDM) are used for this purpose.

Satellite soil moisture data is transformed with an exponential filter to make it comparable to the root zone soil moisture state of the model. The exponential filter formulation explicitly incorporates an autocorrelation component in the rescaled observation, however, the error structure of this operator has been treated until now as an independent Gaussian process. In this work, the variance of the rescaled observation error is estimated based on the residuals from the rescaled satellite soil moisture and the calibrated model soil moisture state. Next, the observation error structure is treated as a Gaussian independent process with time-variant variance; a weakly autocorrelated random process (with autocorrelation coefficient of 0.2) and a strongly autocorrelated random process (with autocorrelation coefficient of 0.8). These experiments are compared with a control case which corresponds to the commonly used assumption of Gaussian independent observation error with time-fixed variance.

Model error is represented by perturbing rainfall forcing data and soil moisture state. These perturbations are assumed to represent all forcing and model structural/parameter errors. Error parameters are calibrated by applying two discharge ensemble verification criteria. Assimilation results are compared and the impacts of the observation error structure assumptions are assessed.

The study area is the semi-arid 42,870 km^2 Warrego at Wyandra River catchment, located in Queensland, Australia. This catchment is chosen for its flooding history, along with having geographical and climatological conditions that enable soil moisture satellite retrievals to have higher accuracy than in other areas. These conditions include large area, semi-arid climate and low vegetation cover. Moreover, the catchment is poorly instrumented, thus satellite data provides valuable information.

Results show a consistent improvement of the model forecast accuracy of the control case and in all experiments. However, given that a stochastic assimilation is designed to correct stochastic errors, the systematic errors in model prediction (probably due to the inaccurate forcing data within the catchment) are not addressed by these experiments. The assumed observation error structures tested in the different experiments do not exhibit significant effect in the assimilation results. This case study provides useful insight into the assimilation of satellite soil moisture retrievals in poorly instrumented semi-arid catchments.

Keywords: *Data assimilation, soil moisture, satellite retrievals, rainfall-runoff model, hydrology*

1 INTRODUCTION

Accurate soil moisture predictions can lead to better modelling of hydrological processes including runoff, groundwater recharge and evapotranspiration. For example, it was shown that runoff prediction could be improved by assimilating antecedent soil moisture into rainfall-runoff modelling (Crow et al., 2005). Nonetheless, the improvement in model skill resulting from assimilating soil moisture observations (on-site or remotely sensed) into rainfall-runoff models has not been fully assessed due to three main limitations: observation uncertainties, temporal resolution and the spatial mismatch between observations and soil water content from rainfall-runoff models (Brocca et al., 2012).

Ground measurements of soil moisture are scarce in most regions, which positions satellite retrievals as a potential solution for improving soil moisture representation. However, satellite soil moisture retrievals have in general higher uncertainty than ground measurements, a coarser spatial resolution and represent only the top few centimetres of soil, all factors that need to be accounted for their use. Exploring tools for improving runoff predictions using satellite soil moisture retrievals has become very popular. The success of stochastic assimilation relies on several factors including whether soil moisture is a dominant control on the runoff generation process in the catchment, the representativeness and accuracy of the observations, and having an adequate representation of model and observation errors. It has been shown that incorrect assumptions regarding the model structure and observation errors can degrade the performance of the stochastic data assimilation (Crow and van Loon, 2006; Crow and Reichle, 2008; Crow and van den Berg, 2010; Reichle et al., 2008; Ryu et al., 2009). Up to date, for hydrologic applications, the error structure of these observations has not been carefully investigated and their assimilation into rainfall-runoff models has been undertaken using the observed time series (i.e., one single realisation from stochastic process) and assuming a time invariant error variance.

In this study, we show how observation error structure assumptions affect the improvement in assimilation skill using a soil moisture product from the Advance Microwave Scanning Radiometer (AMSR-E) and the probability distributed model (PDM). Additionally, we treat the observation as a stochastic process represented by a Monte Carlo - based ensemble. For this we set up an ensemble Kalman filter (EnKF) scheme. The depth mismatch between observed soil moisture (few centimetres of soil) and the predicted soil moisture (depth depending on the calibrated model parameters, but more comparable to the root zone layer) is addressed by applying an exponential filter to the surface observations (Wagner et al., 1999). This filter transforms the surface soil moisture into a profile soil moisture through the estimation of a soil wetness index (SWI). Subsequently, systematic differences between the SWI and the predicted soil moisture are removed by a linear regression rescaling. In different assimilation experiments, the observation error structure is treated as a sequentially independent Gaussian process or as an autocorrelated random process. For evaluating the impacts of these assumptions in the assimilation results, one control case and 3 experiments are defined. The control case corresponds to the commonly used assumption of time invariant variance of the observation error and the assimilation of the observed time series (single realisation). Experiment 1 assumes Gaussian independent error in the observation and treats the observation as a stochastic process so the assimilation is made based on an ensemble of observations. Experiments 2 and 3 assume a “weakly” autocorrelated observation error (Exp.2) and a “strongly” autocorrelated observation error (Exp.3). The assimilation of these two last experiments uses the observed time series (single realisation). Assimilation results are compared and evaluated for the different observation error structure assumptions.

2 STUDY AREA AND DATA

The study area is the Warrego River catchment (42,870 km^2), located in south west Queensland, Australia (see Fig.1). The mean annual precipitation over the catchment is 520 mm and it has a long history of flooding, with at least 10 major events in the last 100 years that have caused extensive inundation of towns and rural lands (*http://www.bom.gov.au/qld/flood/brochures/warrego*).

Rainfall data was obtained from the Australian Water Availability Project (AWAP), which covers the period from 1900-up to date and has a spatial resolution of 0.05° (Jones et al., 2009). Hourly streamflow records for Warrego at Wyandra gauge were collected from the Queensland Department of Natural Resources and Mines website (*http://watermonitoring.derm.qld.gov.au*) for 1967-2013 period. The soil moisture dataset was obtained from the Advance Microwave Scanning Radiometer (AMSR-E), version 5 C/X-band, 0.25° resolution level 3 product for the period 07/2002-10/2011 (Owe et al., 2008).

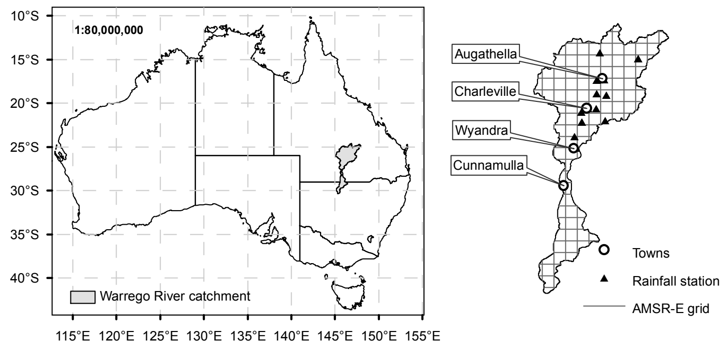

Figure 1. Warrego river catchment

3 METHODS

3.1 Rainfall-runoff model

The probability distributed model (PDM) is a conceptual rainfall-runoff model that has been widely used in hydrologic research (Moore, 2007). The model treats soil water content as a probability distributed variable. Then, two cascade reservoirs are used for representing surface storage and one routing reservoir for representing sub-surface runoff generation. The main inputs of the model are rainfall and potential evaporation. A detailed description of the model structure and formulations is presented by Moore (2007). Here, a lumped model of the catchment and a daily time step is used. The model is calibrated by using a genetic algorithm (Chaturvedi, 2010) with an objective function based on the Nash-Sutcliffe statistic.

3.2 EnKF formulation

The Kalman filter is a Bayesian estimator that sequentially updates model background predictions with available observations. The updating step is based on the relative values of the uncertainties (error covariance) existing in the model and the observations. In the ensemble Kalman filter (EnKF), the error covariance is explicitly calculated from Monte Carlo-based ensembles. For a state-updating assimilation approach, the state ensemble is created by perturbing forcing data and/or the state of the model with unbiased errors.

Let $\theta^-(t) = \{\theta^-_{1,t}, \theta^-_{2,t}, ..., \theta^-_{N,t}\}$ be the perturbed model soil moisture state ensemble prediction (background prediction) before the updating step for time step t, where N is the number of ensemble members. Given that there is no knowledge of the real state values, the ensemble average is use as reference to estimate the prediction error. The error of member i ($\theta^{-\prime}_{i,t}$), the abnormality matrix of the ensemble ($\theta^-_M(t)$), and the covariance matrix of the state model errors (P^-_t), for each time step t, are calculated by:

$$\theta^{-\prime}_{i,t} = \theta^-_{i,t} - \frac{1}{N}\sum_{i=1}^{N}\theta^-_{i,t} \quad ; \quad \theta^-_M(t)' = \{\theta^{-\prime}_{1,t}, \theta^{-\prime}_{2,t}, ..., \theta^{-\prime}_{N,t}\} \quad ; \quad P^-_t = \frac{1}{N-1}\theta^-_M(t)' \times \left(\theta^-_M(t)'\right)^T \tag{1}$$

When a soil moisture observation is available, the SWI is estimated and rescaled (see Section 3.3), and each member of the state ensemble is updated by the rescaled observation, $\theta_{obs(EF)}(t)$, using the following expression:

$$\theta^+_{i,t} = \theta^-_{i,t} + K(\theta_{obs(EF)}(t) - H(\theta^-_{i,t})) \qquad \text{with} \qquad K = \frac{P^-_t H^T}{HP^-_t H^T + R_t} \tag{2}$$

where K is the Kalman gain, H is the observation operator that relates the modelled state to the measured variable. As the observation is rescaled separately prior to the state updating, H reduces to the identity matrix in this work. R_t is the error variance of the rescaled observation for time t.

3.3 Satellite soil moisture rescaling and observation error estimation

Satellite soil moisture retrievals (θ_{obs}) represent the top few centimetres of the soil, while the rainfall-runoff model soil moisture state accounts for a significantly deeper layer. The depth of the modelled storage depends on the calibrated model parameters, but typically it is comparable with the root zone soil moisture. For transferring θ_{obs} information into the soil water content space of the model (θ), we use the exponential filter proposed by Wagner et al. (1999). This filter assumes that the variation in time of the root zone soil moisture is linearly related to the difference between surface soil moisture and root zone soil moisture. The filter estimates a profile average saturation degree by recursively calculating a soil wetness index (SWI) every time there is a satellite soil moisture retrieval θ_{obs} (Brocca et al., 2010):

$$SWI(t) = SWI(t-1) + G_t\left[\theta_{obs}(t) - SWI(t-1)\right] \qquad \text{with} \qquad G_t = \frac{G_{t-1}}{G_{t-1} + e^{-\left(\frac{t-(t-1)}{T}\right)}} \tag{3}$$

G_t is a gain term varying between 0 and 1. T is a calibrated parameter representing the time scale of the SWI variation. SWI is then linearly rescaled in order to meet the same mean and standard deviation as the root zone soil moisture from the model (θ). The rescaled observation is named $\theta_{obs(EF)}$.

The estimation of observation uncertainties is a major challenge, especially given the lack of ground measurements of soil moisture in most areas. In this study we propose to determine an upper boundary of the rescaled observation error variance. If we assume that the error is independent of the measurement (i.e., orthogonal), we can express the variance of the rescaled observation as $Var(\theta'_{obs(EF)}) = Var(\theta_{obs(EF)}) + R$, where R is the rescaled observation error variance from Eq. 2 and $Var(\theta'_{obs(EF)})$ is directly calculated from the rescaled data. Given that the variance is always positive, we can use $Var(\theta'_{obs(EF)})$ as the upper boundary of R. In a first stage, R is considered to be equal to the upper boundary.

Once we have the variance of the rescaled soil moisture error (R), the following experiments are defined in order to explore how error structure assumptions affect the assimilation results:

- Control case: error is treated as a fixed, time invariant variance.
- Exp.1: error is treated as white Gaussian process lacking auto-correlation.
- Exp.2: error is treated as a "weakly" autocorrelated process, with lag-1 day coefficient AR(1)=0.2.
- Exp.3: error is treated as a "strongly" autocorrelated process, with lag-1 day coefficient AR(1)=0.8.

The random component of the autocorrelated errors in Exp.2 and Exp.3 is assumed to be zero mean autocorrelated Gaussian noise, the spread of which is calculated by constraining the temporal mean of the rescaled observation ensemble covariance to be equal to R.

3.4 Model error estimation

Model error is represented by perturbing forcing precipitation data with an independent multiplicative log-normally distributed error (mean 1 and standard deviation σ_p), and by perturbing the soil moisture with independent additive normally distributed error (mean 0 and standard deviation σ_{sm}). These perturbations consider input forcing, model parameter and model structure error sources. Model error parameters (σ_p and σ_{sm}) are calibrated by running the open-loop (perturbing forcing and soil moisture state, but without assimilating rescaled soil moisture observations) for 100 ensemble members (N), and evaluating the following two discharge ensemble verification criteria:

i) If the ensemble spread is large enough, the temporal average of the ensemble skill (sk_t) should be similar to the temporal average of the ensemble spread (sp_t), i.e., $\overline{sk}/\overline{sp} = 1$ (Brocca et al., 2012; De Lannoy et al., 2006), where:

$$\overline{sk} = \frac{1}{T}\sum_{t=1}^{T}\left[\overline{Q_{sim}}(t) - Q_{obs}(t)\right]^2 \qquad \text{and} \qquad \overline{sp} = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{N}\sum_{i=1}^{N}\left(Q_{sim}(i,t) - \overline{Q_{sim}}(t)\right)^2\right] \tag{4}$$

ii) If the observation is indistinguishable from a member of the ensemble, the ratio between $\overline{sk}$ and the ensemble mean-square-error ($\overline{mse}$) should be equal to $\sqrt{N+1/2N}$ (Moradkhani et al., 2005; Brocca et al., 2012), where:

$$\overline{mse} = \frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{N}\sum_{i=1}^{N}\left(Q_{sim}(i,t) - Q_{obs}(t)\right)^2\right] \tag{5}$$

4 RESULTS AND DISCUSSION

4.1 Model calibration

Figure 2 presents the simulated and observed discharge time series for both the calibration and verification periods. These results reveal that calibrated model underestimates the observed peak flows and overestimates low flows. A likely factor that contributes to the performance of the model is the poor density of rainfall gauges within the catchment, which results in low quality gridded rainfall data for the area. Model performance can also be related to the objective function used for calibration (maximising the Nash-Sutcliffe efficiency) (Gupta and Kling, 2011). Given the semi-arid nature of the catchment, rainfall-runoff generation processes

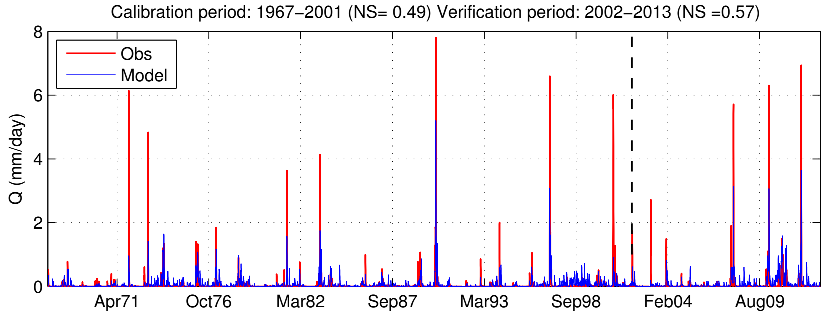

Figure 2. Discharge prediction time series, the dashed black line indicates the end of calibration period.

are likely to be dominated by the soil water content of the catchment, which would explain why many rainfall events do not result in discharge (this can be seen by comparing runoff ratios with rainfall intensity and with modelled soil moisture, not shown here). Presumably, given the poor representation of the forcing data, the model structure and conceptualisation are not able to correctly represent this saturation excess runoff process. Due to the lack of ground data for improving rainfall representation over the catchment, and the likely high dependency of runoff generation processes on the soil water content of the catchment, the assimilation of satellite soil moisture offers an important opportunity for improving the model discharge prediction.

4.2 Model and observation error estimation

The linear rescaling described in Section 3.3 was trained using the first two years of data (2002-2004) and then updated in each time step. The parameter T of the exponential filter (eq. 3) was calibrated for the same training window. The rescaled observation time series is presented in Figure 3, and has a correlation coefficient of 0.82 with the modelled soil moisture for the verification period. The standard deviation of the associated residuals (std_{res}) is 0.05 m^3/m^3 (expressed as volumetric percentage of the calibrated soil moisture storage of 630 mm). These results reveal the strong concordance between the model soil moisture state and the rescaled surface soil moisture observation. The observation error variance is estimated as 1360 mm^3. The adopted standard deviation of the random component of the "weak" and "strong" autoregressive processes (exp. 2 and 3) are 0.0574 m^3/m^3 and 0.0351 m^3/m^3, respectively. These values fullfill the constraint of a temporal observation variance equals to R. Following the methodology described in section 3.4, model error parameters calibration results in σ_P = 0.3195 and σ_{sm} = 0.0248 m^3/m^3 (volumetric percentage of soil moisture storage).

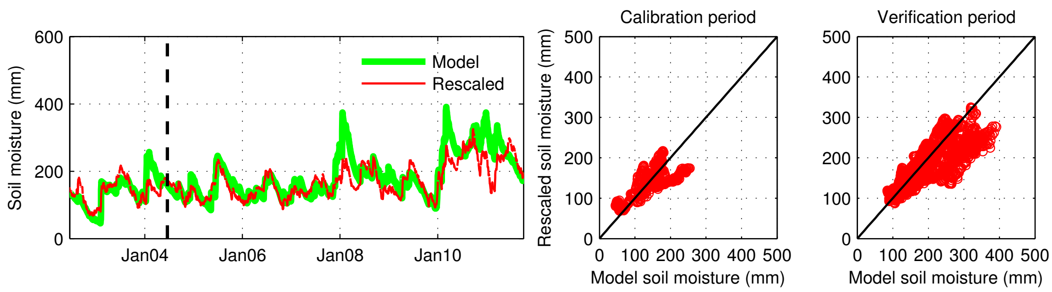

Figure 3. Rescaled observations, dashed black line indicates the end of training period.

4.3 Assimilation experiments

The evaluation of assimilation is undertaken for the period 06/2004-10/2011. The first half of the analysis window, up to 03/2008, is characterised by small flow events (Period 1) while the second half (03/2008-10/2011) is characterised by larger flow events, having at least three major flood events (Period 2).

The assimilation results for experiment 3 are presented in Fig. 4, for these two separate periods. The green dashed line represents the un-perturbed model (i.e. the predictions of the calibrated model, called "sim"). From these graphs it can be seen that for small flow events (Period 1), the assimilation procedure is reducing the

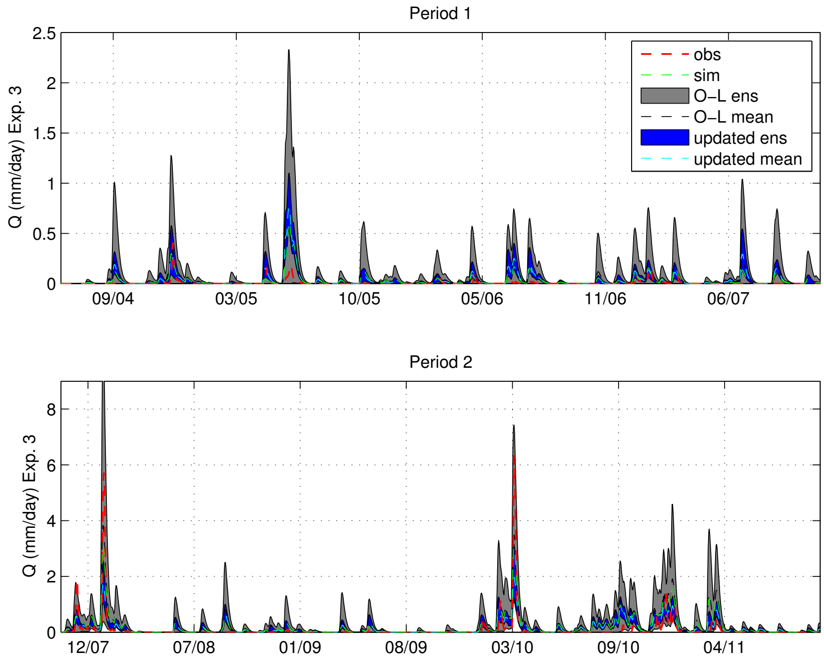

Figure 4. Assimilation results for experiment 3 (observation error treated as a strong autocorrelated process, with lag-1 day coefficient AR(1)=0.8).

Table 1. Evaluation metrics of assimilation results for control case and experiments

Experiment	RMSD sim		MRMSD open loop		MRMSD updated		NRMSD	
	Period 1	Period 2	Period 1 (0.01)*	Period 2 (0.02)*	Period 1 (0.002)*	Period 2 (0.005)*	Period 1 (0.12)*	Period 2 (0.06) *
Control case	0.05	0.32	0.11	0.44	0.07	0.36	0.65	0.80
Exp.1	0.05	0.32	0.11	0.45	0.06	0.34	0.56	0.76
Exp.2	0.05	0.32	0.12	0.44	0.08	0.36	0.65	0.82
Exp.3	0.05	0.32	0.12	0.47	0.08	0.36	0.65	0.77

* 95% confidence interval

open-loop spread and in general reducing the model overestimation of streamflow (when analysing the mean of the open-loop and updated ensemble, compared with the observed discharge). For larger flow events events (Period 2), the assimilation is mainly reducing the spread of the open-loop while the model underestimation is not being corrected. The assimilation results of the control case and experiments 1 and 2 show similar relation between the open-loop and updated ensembles (not shown here) and their evaluation metrics are summarised in Table 1.

Table 1 presents the mean root mean square difference (MRMSD) and the normalised MRMSD (NRMSD) for the different assimilation experiments for Periods 1 and 2. The NRMSD is calculated as the ratio between the MRMSD from the open loop ensemble and the MRMSD from the updated ensemble. Additionally, the RMSD of unperturbed model (sim) is presented in the Table. Data assimilation for the control case and the experimental cases results in an improvement of around 40% in period 1 and of 20% in period 2, in terms of MRMSD. Differences between the control case and the experiments are within the confidence intervals thus they are not considered significant.

In general, the spread of the discharge ensemble is reduced by assimilating satellite soil moisture retrievals, but the poor representation of the model, evaluated as the ensemble mean compared with the observed discharge, is not consistently corrected (it only improves for specific events). Given that stochastic assimilation is designed to correct stochastic errors, the model systematic errors (presumably coming from the poor representation of precipitation over the catchment, given the lack of instrumentation within the area) are not addressed thus the performance of the assimilation becomes marginal. Moreover, the different observation error structures tested does not affect the assimilation results. This suggests that even though observation error structure theoretically has a direct effect on an EnKF-based assimilation, when working with real data and the uncertainties inherent in a poorly instrumented area, the effect is trivial.

5 CONCLUSIONS

This work has shown that the assimilation of satellite soil moisture retrievals derived from AMSR-E into PDM results in a consistent improvement of the model predictions. This improvement is based on the reduction of the model forecast uncertainty. Nevertheless, given that a stochastic assimilation is designed to correct stochastic errors (which translates in the achieved reduction of model forecast uncertainty), the systematic poor model performance (probably due to poor representation of forcing data within the catchment) is not addressed by these experiments. While the spread of the ensemble discharge prediction is reduced after assimilation, the ensemble mean is not always closer to the discharge observation. Moreover, the different observation error structures tested here did not result in significant differences in the assimilation performance. This suggests that when the model prediction accuracy and uncertainties are mainly controlled by high uncertainties in forcing data, the assumptions of the observation error structure made in a state-update assimilation framework have little effect. These findings enhance our understanding of the advantages and limitations of assimilating satellite soil moisture observations into a rainfall-runoff model for improving streamflow prediction. In order to address the systematic model predictions biases, while reducing the stochastic errors of the model, efforts should be focused on combining the presented state-update assimilation scheme with some tool to reduce the uncertainty in rainfall data.

ACKNOWLEDGEMENT

This research was conducted with financial support from the Australian Research Council (ARC Linkage Project No. LP110200520) and the Bureau of Meteorology, Australia. We gratefully acknowledge the advise and data provision of Chris Leahy and Soori Sooriyakumaran from the Bureau of Meteorology, Australia.

REFERENCES

Brocca, L., F. Melone, T. Moramarco, W. Wagner, V. Naeimi, Z. Bartalis, and S. Hasenauer (2010). Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrology and Earth System Sciences 14*(10), 1881–1893.

Brocca, L., T. Moramarco, F. Melone, W. Wagner, S. Hasenauer, and S. Hahn (2012). Assimilation of surface- and root-zone ASCAT soil moisture products into rainfall–runoff modeling. *Geoscience and Remote Sensing, IEEE Transactions on 50*(7), 2542–2555.

Chaturvedi, D. (2010). Matlab Program of Genetic Algorithms.

Crow, W. T., R. Bindlish, and T. J. Jackson (2005). The added value of spaceborne passive microwave soil moisture retrievals for forecasting rainfall-runoff partitioning. *Geophysical Research Letters 32*, L18401.

Crow, W. T. and R. H. Reichle (2008). Comparison of adaptive filtering techniques for land surface data assimilation. *Water Resources Research 44*, W08423–.

Crow, W. T. and M. J. van den Berg (2010). An improved approach for estimating observation and model error parameters in soil moisture data assimilation (doi 10.1029/2010WR009402). *Water Resources Research 46*, W12519.

Crow, W. T. and E. van Loon (2006). Impact of Incorrect Model Error Assumptions on the Sequential Assimilation of Remotely Sensed Surface Soil Moisture. *Journal of Hydrometeorology 7*, 421–432.

De Lannoy, G. J., P. R. Houser, V. Pauwels, and N. E. Verhoest (2006). Assessment of model uncertainty for soil moisture through ensemble verification. *Journal of Geophysical Research: Atmospheres (1984–2012) 111*(D10).

Gupta, H. V. and H. Kling (2011). On typical range, sensitivity, and normalization of mean squared error and Nash-Sutcliffe efficiency type metrics. *Water Resources Research 47*(10), W10601.

Jones, D. A., W. Wang, and R. Fawcett (2009). High-quality spatial climate data-sets for australia. *Australian Meteorological and Oceanographic Journal 58*(4), 233.

Moore, R. J. (2007). The PDM rainfall-runoff model. *Hydrology & Earth System Sciences 11*(1), 483–499.

Moradkhani, H., S. Sorooshian, H. Gupta, and P. Houser (2005). Dual state–parameter estimation of hydrological models using ensemble Kalman filter. *Advances in Water Resources 28*(2), 135–147.

Owe, M., R. de Jeu, and T. Holmes (2008). Multisensor historical climatology of satellite-derived global land surface moisture. *Journal of Geophysical Research: Earth Surface (2003–2012) 113*(F1).

Reichle, R. H., W. T. Crow, and C. L. Keppenne (2008). An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resources Research 44*(3).

Ryu, D., W. T. Crow, X. Zhan, and T. J. Jackson (2009). Correcting Unintended Perturbation Biases in Hydrologic Data Assimilation. *Journal of Hydrometeorology 10*, 734–750.

Wagner, W., G. Lemoine, and H. Rott (1999). A method for estimating soil moisture from ERS scatterometer and soil data. *Remote Sensing of Environment 70*(2), 191–207.