

Joint model state-parameter retrieval through the evolutionary data assimilation approach

Gift Dumedah^a, and Jeffrey P. Walker^a

^a*Department of Civil Engineering, Monash University, Melbourne, VIC, Australia, 6845
Email: gift.dumedah@monash.edu.au*

Abstract: Data assimilation (DA) methods have been widely used to improve model state estimation by merging model outputs with observations. Traditionally, studies have focused on updating model state variables but recent studies have augmented model parameters alongside model state variables to improve the estimation procedure. The updated model ensemble members represent a compromised estimation between prediction and observation. The compromise, which is usually in objective space subject to agreement between observation and model predictions, is important. However, few studies have actually employed DA procedures to investigate the updated members in decision space, through examination of the temporal changes of model states and parameters. Usually, the model states and parameters evolve/change: (i) subject to changes in observation, (ii) to account for the varied uncertainties in different land surface conditions, and (iii) due to their intricate connection with hydrologic conditions which evolve across assimilation time periods. Moreover, the update procedure in most DA methods is controlled predominantly by matchings between observation and model predictions with limited impact from decision space through model state variables and parameters. As a result, DA procedures are needed to tightly link the compromise in objective space to decision space, with the capability to examine the temporal changes of model states and parameters.

To address these challenges, this study has employed the Evolutionary Data Assimilation (EDA) method in a joint state-parameter estimation to assimilate: (i) synthetic daily soil moisture into the Joint UK Land Environment Simulator (JULES), and (ii) hourly streamflow into the Hydrologiska Byråns Vattenbalansavdelning (HBV) model. The EDA is a relatively new formulation of the multi-objective evolutionary strategy for the purpose of data assimilation. The EDA was applied to illustrate its capability to both retrieve model parameter values and to improve estimation of soil moisture and streamflow in a two-step procedure.

In the soil moisture assimilation, the first step involved the generation of initialization model state and parameter values as the original ‘truth’ where they were applied into the JULES model to simulate surface soil moisture, representing the synthetic soil moisture. The second step assimilated the synthetic soil moisture into a perturbed version of the JULES model to retrieve the original model state and parameter values. The updated model states and parameter values were compared to the original ‘truth’ to show that the EDA can both retrieve the original ‘truth’ parameter values and the soil moisture states. The soil moisture assimilation was illustrated for the Yanco region in New South Wales, Australia from January to December 2010.

In the streamflow assimilation, the original ‘truth’ values of the HBV model were obtained from two independent studies generated using: (i) the Multistart weight-adaptive recursive parameter estimation, and (ii) the Shuffled Complex Evolution. The EDA was applied to assimilate hourly streamflow into the HBV model to retrieve the calibrated model state and parameter values. The streamflow assimilation was illustrated for the Bellebeek catchment in Belgium from August 2006 to July 2007. The input data for the HBV model and data sets in the Bellebeek catchment were provided by the authors of the two independent studies.

The synthetic soil moisture were compared to the EDA updated soil moisture in the soil moisture assimilation, whereas the observation streamflow were compared to the EDA updated streamflow. The findings show a high estimation accuracy of the EDA for soil moisture and streamflow based on the evaluation measures and the two independent studies. Moreover, the updated ensemble of model states and parameter values were evaluated across the assimilation time steps showing the level of convergence for model state variables and parameters. The evaluation of the temporal evolution of updated ensemble members in decision space demonstrates the capability of the EDA to retrieve model state variables and parameters. Thus, the key potential of the EDA lies in the evaluation of updated members for model state and parameter linkages in decision space.

Keywords: *Soil moisture, Streamflow, Data assimilation, State-parameter retrieval, Evolutionary strategy.*

1 INTRODUCTION

Hydrologic data assimilation (DA) has increasingly focused on state estimation [Xie and Zhang, 2010; Camporese et al., 2010], with relatively few examples where the model parameters have been augmented alongside model state variables [Vrugt et al., 2005; Moradkhani and Hsu, 2005]. While the focus has been to improve the overall compromise between observation and model estimation, very few studies have actually employed DA procedures to examine the temporal changes for model states and parameters. Typically, the model states and parameters evolve across data assimilation time periods because they are intricately linked to the hydrological conditions of the area under study. Moreover, the temporal changes in model states and parameters are necessary to account for the varied uncertainties in different land surface conditions across different time periods. The lack of evaluation of the temporal changes for model state variables and parameters is partly because the updated ensemble members generated by most DA methods are not usually adjusted for model state variables and parameters. This is because in most DA methods, the updated ensemble members are generated subject to an update procedure which is predominantly a compromise in objective space (subject to observation and prediction) with limited control from decision space (that is, model states and parameters).

The problem of temporal changes for model states and parameters is addressed in this study through the evolutionary data assimilation (EDA) procedure. The EDA is a fairly new DA methodology that employs an advanced analytical evolutionary strategy to merge and account for uncertainties in model prediction and observation data. Evolutionary algorithms have been applied in several hydrological studies including [Dumedah et al., 2012; Ines and Mohanty, 2009; Chemin and Honda, 2006], and the EDA has been applied in assimilating soil moisture [Dumedah and Coulibaly, 2012a; Dumedah et al., 2011], and streamflow in gauged and ungauged catchments [Dumedah, 2012; Dumedah and Coulibaly, 2012b].

This study assesses the capability of the EDA to reproduce known model parameters and states for a land surface model, and a conceptual hydrologic model. Specifically, the EDA estimates model states and parameters in: (i) soil moisture assimilation for the Joint UK Land Environment Simulator (JULES) model, and (ii) streamflow assimilation for the Hydrologiska Byråns Vattenbalansavdelning (HBV) model. The soil moisture assimilation in JULES employed a synthetic study to examine the capability of the EDA to retrieve original ‘truth’ values for model parameters and states. The study is demonstrated for the Yanco area in New South Wales, Australia using synthetic soil moisture data from January to December 2010. The streamflow assimilation in HBV compares the retrieved values for model states and parameters to corresponding values obtained from independent studies in Ferket et al. [2010], and Pauwels and De Lannoy [2009]. The streamflow assimilation is illustrated for the Bellebeek catchment in Belgium from August 2006 to July 2007.

2 MATERIALS AND METHODS

2.1 The evolutionary data assimilation - EDA

The EDA is a formulation of evolutionary algorithm into a data assimilation procedure [Dumedah and Walker, 2013; Dumedah and Coulibaly, 2012a]. Evolutionary algorithms are population based search procedure using the concept of biological evolution and natural selection to address complex problems. In evolutionary strategy, a population of candidate members (or solutions) are allowed to compete based on evaluation conditions, where the high performing members are selected and varied to reproduce new members.

The implementation of the EDA used in this study follows Dumedah and Walker [2013]; Dumedah [2012]. In the EDA procedure (Figure 1), a population of model predictions are generated by perturbing model parameters, states, and input forcing data. A corresponding number of observation ensemble members are generated based on the observation data and its associated uncertainty. The ensemble predictions and observations are used at the evaluation stage to determine the fitness for each member based on: the absolute difference, and the cost function [Dumedah and Walker, 2013]. The fitter members with the optimal compromise between the simulated model output and the observation, are selected and varied to reproduce new members for the population. This complete cycle comprising the generation of population members, evaluation, selection, and reproduction of new members constitutes one cycle of a generation.

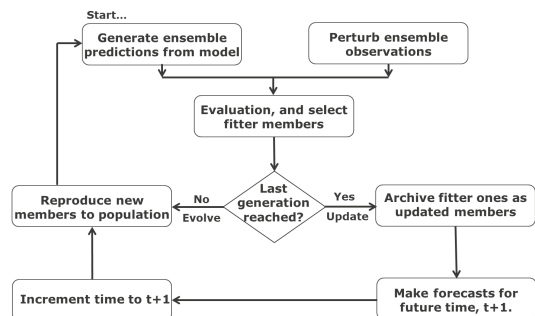


Figure 1: A simplified procedure of the evolutionary data assimilation, adapted from Dumedah [2012].

The population members are evolved across several generations to determine the final evolved members which represent the updated members. For each assimilation time step, the evolution is repeated to determine the updated members.

2.2 Description of the Yanco area, the JULES model, data sets, and the assimilation setup

The Yanco area is a 60×60 -km land area located in the south east Australia with latitude ranging from 34.5 to $35.2^\circ S$, and longitude between 145.0 and $146.8^\circ E$. The landscape is predominantly flat with sandy loam textured soils along with scattered clays and sand. The vegetation is mainly rain-fed cropping/pasture with scattered trees and grassland, with some irrigated crops in the north west.

The JULES model is a widely used tiled model of sub-grid heterogeneity which simulates water and energy fluxes between a vertical profile of soil layers, vegetation, and the atmosphere [Best et al., 2011]. The JULES model uses input forcing data, surface land cover data, soil properties data, and values for prognostic variables. To set up the JULES model for the Yanco area, the soil properties data were derived from the Digital Atlas of Australian Soils [McKenzie et al., 2000], obtained through the Australian Soil Resource Information System. The soil data include information on soil texture class, along with clay content, bulk density, saturated hydraulic conductivity, and soil layer thickness for horizons A and B [McKenzie et al., 2000]. The information for surface land cover was obtained from the Australian National Dynamic Land Cover Dataset [Lymburner et al., 2011], derived from the 250-m bands of the Moderate Resolution Imaging Spectroradiometer (MODIS). The input forcing data were obtained from the Australian Community Climate Earth-System Simulator (ACCESS-A) providing an hourly data with 12-km spatial resolution [Bureau of Meteorology, 2010]. The ACCESS-A precipitation data set was bias corrected using a 5-km gridded rain gauge precipitation data from the Australian Water Availability Project [Jones et al., 2009].

To set up the synthetic experiment, initial model states and parameters were randomly generated using the lower and upper bounds of the model state variables and parameters as a constraint. These model states and parameter values (representing the original ‘truth’ values) were then applied into the JULES model to generate synthetic soil moisture truth data from January to December 2010. The EDA was then applied with time-variant model state and parameter values using the synthetic truth soil moisture values as observation data in a daily time step assimilation into the JULES model. In the EDA run, the initial population of ensemble members was generated using the lower and upper bounds of the model state variables and parameters. The variance of an ensemble soil moisture was used as the observation uncertainty, and the time-variant model error was estimated from the population of ensemble members. The assimilation was run with an ensemble size of 200 at a daily time step from January to December 2010. In the EDA approach, a subset of 20 members was selected from the entire 200 members for each assimilation time step, to represent the updated ensemble members. Note that the updated members comprise model states and parameter values, with their associated estimates of soil moisture. The updated soil moisture from the EDA was compared to the synthetic soil moisture observation, and the updated model state and parameter values from the EDA were compared to the original truth values in the results section in 3.1.

2.3 Description of the Bellebeek catchment, the HBV model, data sets, and the assimilation setup

The Bellebeek is a catchment located in Belgium with a drainage area of about 86.36 -km² at a latitude of $50.6^\circ N$, and a longitude of $4.1^\circ E$. The landscape is varied with elevation ranging from 10 to 110-m, along with mainly loam (about 75%) textured soils, and land use is predominantly agriculture (about 63%) with scattered pasture (about 22%).

The HBV model is a conceptual rainfall-runoff model with input forcing data including precipitation, air temperature and evapotranspiration. The HBV model was used to simulate the streamflow in the Bellebeek catchment. The HBV model has been calibrated for streamflow at an hourly time step from 1 August 2006 through 31 July 2007 for the Bellebeek catchment by Ferket et al. [2010], and Pauwels and De Lannoy [2009] using two calibration methods: the Multistart weight-adaptive recursive parameter estimation (MWARPE) method, and the Shuffled Complex Evolution (SCE-UA).

In the EDA, the HBV model and its model parameter and state intervals were applied to assimilate streamflow at an hourly time step. An ensemble size of 400 was used with a subset of 20 members selected as the updated ensemble for each assimilation time step, as with the JULES experiment. The larger ensemble size of 400 was used because the computational demand in the HBV was lower compared to the JULES model. The EDA outputs including the estimated values for model states and parameters and the updated flow estimates are compared to the outputs from Ferket et al. [2010], and Pauwels and De Lannoy [2009].

3 RESULTS AND DISCUSSIONS

3.1 Retrieval of synthetic model state variables and parameters in the JULES model

The comparison between the synthetic truth and the EDA estimated soil moisture is shown in Figure 2. This result shows a high agreement between the synthetic truth and the estimated soil moisture values, as evidenced by the accurate values for evaluation measures including root mean square error (RMSE) and coefficient of determination (R^2).

The temporal changes for model states and parameters across assimilation time steps as compared to the original truth values is shown in Figure 3. The estimated values overlap with the original truth values across all model state variables and parameters, indicating that for each assimilation time step the original truth values were successfully retrieved as a member of the updated ensemble members.

To examine the probability distribution of the estimated values, a plot of the probability score and deviation from the mean is shown in Figure 4. The probability distributions show the deviation of the original truth values in relation to the mean and its corresponding probability score. The original truth values are not all located closest to the mean, but the spread of the estimated values covers all the original truth values. It is noted that the probability score values do not indicate the quality of the estimated values, because in the EDA approach each updated ensemble member is equally competitive to providing a unique compromise between the simulated output and the observation.

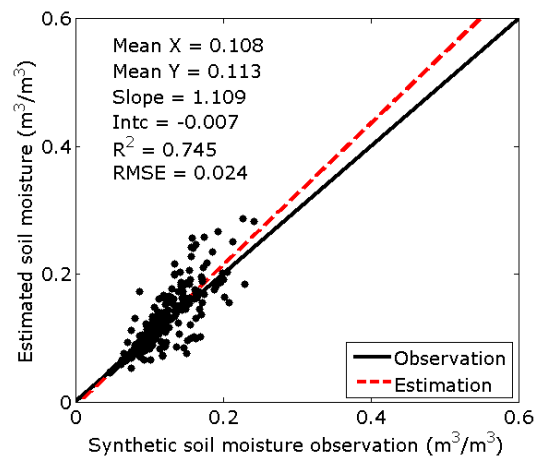


Figure 2: Comparison between the synthetic truth soil moisture and the estimated soil moisture using the EDA

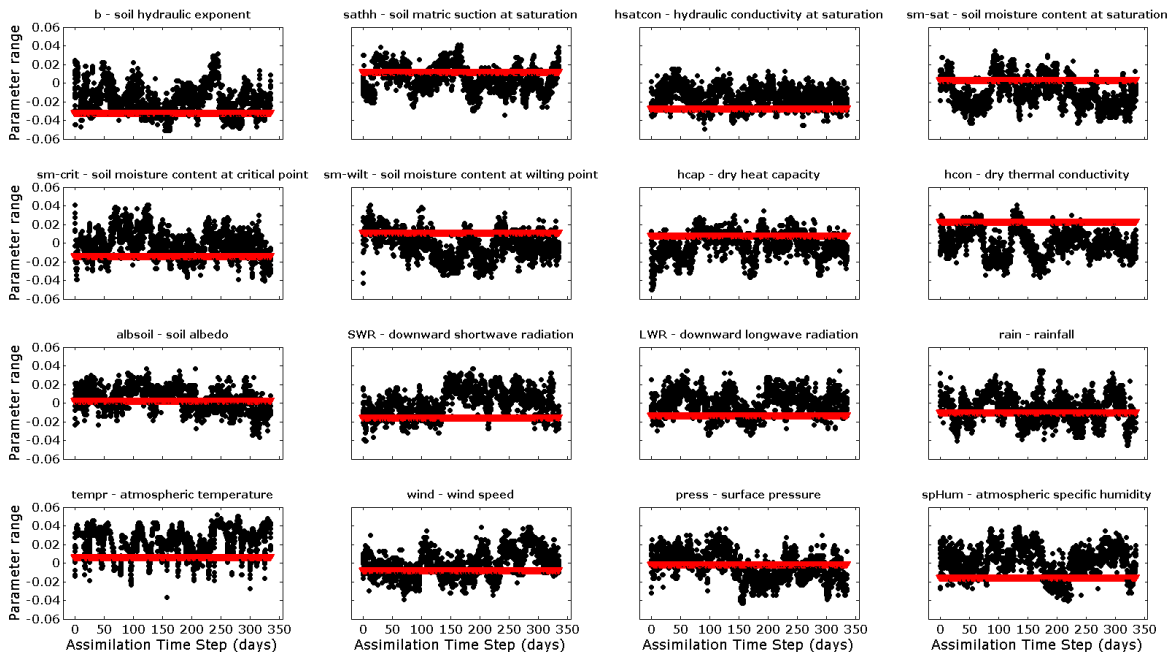


Figure 3: Comparison between the ‘truth’ model state and parameter values and those estimated from the EDA. The red-coloured symbol (i.e., triangle) represents the original synthetic values, and the dark-coloured symbol (i.e., circle) represents the EDA values. Although the synthetic parameter values are plotted to be the same across assimilation time steps, they in fact are applicable only to the initial time step and then propagated across subsequent time steps.

These results illustrate the capability of the EDA approach to explore the changes in model state and parameter values in response to observation data. The ensemble model state and parameter values provide crucial information on the distribution (e.g. whether clustered, scattered) of parameter values for further evaluation with the potential to explore model structure, identify model weaknesses, and to study the representativeness

of physical processes in models. Based on the clustering pattern, it is noted that model state variables and parameters with a high degree of clustering are sensitive, whereas values with low levels of clustering show in-sensitive model state variables and parameters.

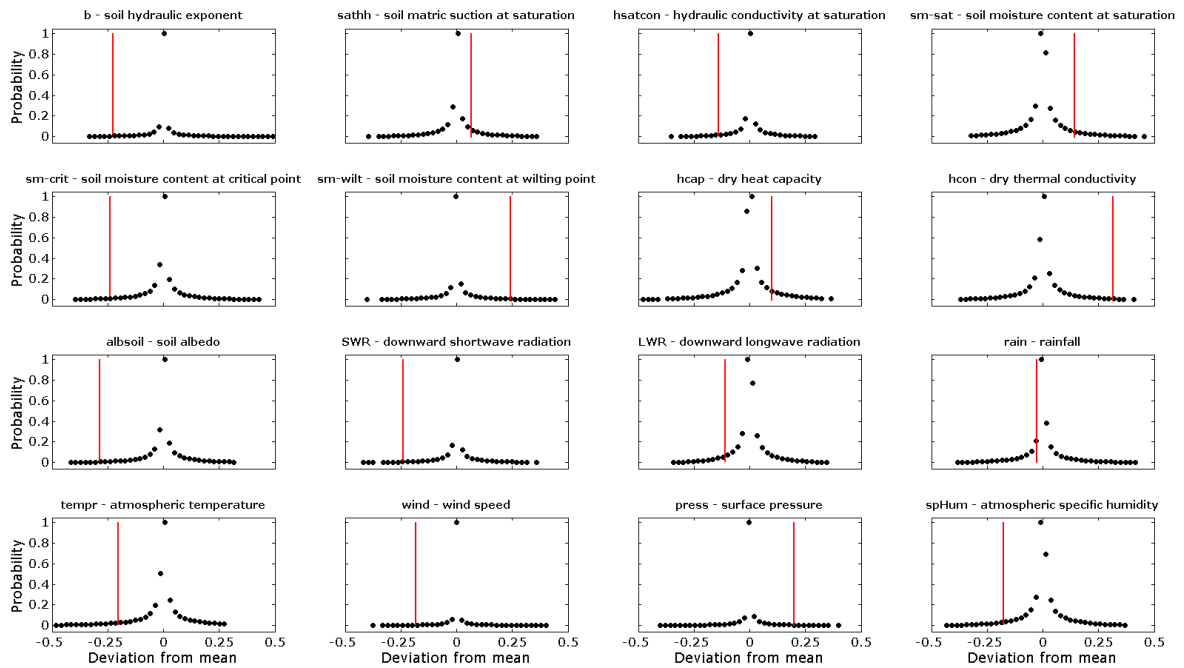


Figure 4: Probability distributions for the updated ensemble values (presented in dark-coloured dots) for model state variables and parameters in comparison to the original truth values (shown in red-coloured lines).

3.2 Retrieval of calibrated model state variables and parameters in the HBV model

The comparison of the observed streamflow to the EDA estimate, and to the calibrated estimate using MWARPE and SCE-UA from Ferket et al. [2010], and Pauwels and De Lannoy [2009] are shown in Figure 5. The high agreement between the observation and the EDA estimated streamflow is supported by the accurate values for RMSE and R^2 , thus providing an improved estimation compared to the MWARPE and SCE-UA estimates.

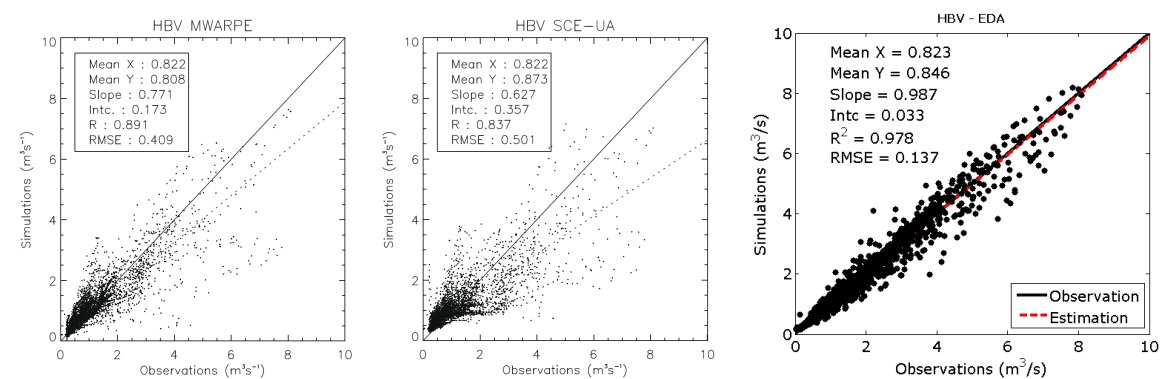


Figure 5: Results of the estimated streamflow in comparison to the observed streamflow from the MWARPE, SCE-UA, and EDA methods. The MWARPE and SCE-UA estimates were obtained from Ferket et al. [2010], and Pauwels and De Lannoy [2009].

The estimated model parameter values across the assimilation time steps are compared to the calibrated values from the MWARPE and the SCE-UA methods in Figure 6. As shown in these results, the calibrated values from MWARPE and the SCE-UA methods overlap with the estimated model parameter values from the EDA. The estimated EDA values show the changes in model parameter values in response to the observation data. The overlap between the estimated EDA values and the calibrated values show the capability of the EDA to both retrieve the calibrated values and the distribution of model parameter values.

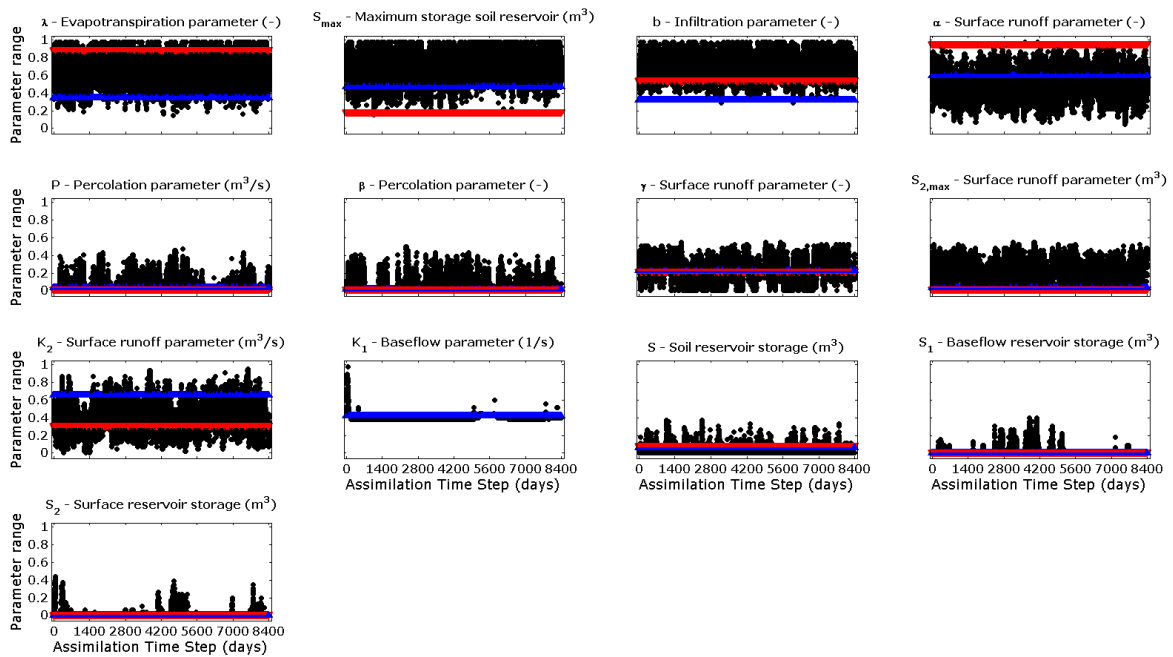


Figure 6: Comparison of model state variables and parameters between estimates from Ferket et al. [2010], and Pauwels and De Lannoy [2009] to those obtained from the EDA. The red-coloured symbol represents the MWARPE values; the blue-coloured symbol shows the SCE-UA values; and the dark-coloured symbol represents the EDA values.

4 FINDINGS AND CONCLUSION

This study has employed the EDA to examine the temporal changes for model states and parameters for soil moisture assimilation into the JULES model, and streamflow assimilation into the HBV model. The EDA estimation of soil moisture in the JULES model and streamflow in the HBV model has been shown to be highly accurate based on the evaluation measures and its comparison to independent outputs from Ferket et al. [2010], and Pauwels and De Lannoy [2009].

The improved estimation of the EDA procedure is accompanied by updated ensemble values for model states and parameters providing the capability to examine their temporal changes subject to changes in observation data. The EDA has been demonstrated to retrieve model states and parameters during DA for soil moisture into the JULES model and for streamflow into the HBV model. The accurate level of model state variable and parameter retrieval for these different data types show the functional capability of the EDA procedure to assimilate different data types into different simulation models. The EDA estimated values illustrate the capability to explore their clustering patterns with the potential to quantify the level of convergence for model state variables and parameters. Moreover, the level of clustering or the persistence of model state variables and parameters which have been subjected to changes in observation data across different assimilation time periods is indicative of the degree of robustness/sensitivity. The presented EDA results for the JULES model have identified sensitive and insensitive model state variables and parameters. It is noted that the key prospects of the EDA approach lie in the evaluation of the updated ensemble members in decision space for model state variables and parameters. This has the potential for the identification of model weakness, better understanding and partitioning of physical processes, and improved model predictions.

ACKNOWLEDGMENTS

This work was supported by funding from the Australian Research Council (DP0879212).

REFERENCES

- Best, M. J., M. Pryor, D. B. Clark, G. G. Rooney, R. L. H. Essery, C. B. Mnard, J. M. Edwards, M. A. Hendry, A. Porson, N. Gedney, L. M. Mercado, S. Sitch, E. Blyth, O. Boucher, P. M. Cox, C. S. B. Grimmond, and R. J. Harding (2011). The Joint UK Land Environment Simulator (JULES), Model description - Part 1: Energy and water fluxes. *Geosci. Model Dev. Discuss.* 4, 595–640, doi:10.5194/gmdd-4-595-2011.
- Bureau of Meteorology (2010). Operational implementation of the access numerical weather prediction systems. *NMOC Operations Bulletin* 83, 1–34.

- Camporese, M., C. Paniconi, M. Putti, and S. Orlandini (2010). Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data. *Water Resources Research* 46, doi:10.1029/2008WR007536.
- Chemin, Y. and K. Honda (2006). Spatiotemporal fusion of rice actual evapotranspiration with genetic algorithms and an agrohydrological model. *IEEE Transactions on Geoscience and Remote Sensing* 44(11), 3462–3469.
- Dumedah, G. (2012). Formulation of the evolutionary-based data assimilation, and its practical implementation. *Water Resources Management* 26(13), 1–18, doi:10.1007/s11269-012-0107-0.
- Dumedah, G., A. A. Berg, and M. Wineberg (2011). An integrated framework for a joint assimilation of brightness temperature and soil moisture using the Non-dominated Sorting Genetic Algorithm-II. *Journal of Hydrometeorology* 12(2), 1596–1609, doi: 10.1175/JHM-D-10-05029.1.
- Dumedah, G., A. A. Berg, and M. Wineberg (2012). Pareto-optimality and a search for robustness: choosing solutions with desired properties in objective space and parameter space. *Journal of Hydroinformatics* 14(2), 270–285, doi:10.2166/hydro.2011.120.
- Dumedah, G. and P. Coulibaly (2012a). Evolutionary assimilation of streamflow in distributed hydrologic modeling using in-situ soil moisture data. *Advances in Water Resources* 53, 231–241, doi:10.1016/j.advwatres.2012.07.012.
- Dumedah, G. and P. Coulibaly (2012b). Integration of evolutionary based assimilation into Kalman-type methods for streamflow simulations in ungauged watersheds. *Journal of Hydrology* 475, doi:10.1016/j.jhydrol.2012.10.033, 428–440.
- Dumedah, G. and P. J. Walker (2013). Evaluation of model parameter convergence when using data assimilation in soil moisture estimation. *Journal of Hydrometeorology*, doi:10.1175/JHM-D-12-0175.1.
- Ferket, B. V., B. Samain, and V. R. Pauwels (2010). Internal validation of conceptual rainfallrunoff models using baseflow separation. *Journal of Hydrology* 381(1-2), 158–173, doi:10.1016/j.jhydrol.2009.11.038.
- Ines, A. and B. Mohanty (2009). Near-surface soil moisture assimilation for quantifying effective soil hydraulic properties using genetic algorithms: 2. using airborne remote sensing during sgp97 and smex02. *Water Resources Research* 45, W01408.
- Jones, D. A., W. Wang, and R. Fawcett (2009). High-quality spatial climate data-sets for Australia. *Australian Meteorological and Oceanographic Journal* 58, 233–248.
- Lymburner, L., P. Tan, N. Mueller, R. Thackway, A. Lewis, M. Thankappan, L. Randall, A. Islam, and U. Senarath (2011). The national dynamic land cover dataset (DLCD), record 2011/31, www.ga.gov.au/landcover,. Technical report, Geoscience Australia.
- McKenzie, N., D. Jacquier, L. Ashton, and H. Cresswell (2000). Estimation of soil properties using the atlas of australian soils. Technical report, CSIRO Land and Water Technical Report 11/00. Available from <http://www.clw.csiro.au/publications/technical2000/>.
- Moradkhani, H. and K. Hsu (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research* 41(W05012), doi:10.1029/2004WR003604.
- Pauwels, V. R. N. and G. J. M. De Lannoy (2009). Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space. *Water Resources Research* 45(8), W08428, doi:10.1029/2008WR007590.
- Vrugt, J. A., B. A. Robinson, and V. V. Vesselinov (2005). Improved inverse modeling for flow and transport in subsurface media: Combined parameter and state estimation. *Geophysical Research Letters* 32, doi:10.1029/2005GL023940.
- Xie, X. and D. Zhang (2010). Data assimilation for distributed hydrological catchment modeling via ensemble kalman filter. *Advances in Water Resources* 33(6), 678–690.