Data assimilation and land surface modelling: the role of open source platforms

B.T. Gouweleeuw [a], **L.J. Renzullo** [a], **J. Shoesmith** [b] **and T.H. Raupach** [a]

[a] *CSIRO Land and Water, GPO Box 1666, ACT 2601, Australia*
Email: ben.gouweleeuw@csiro.au
[b] *CSIRO Land and Water, GPO Box 2583, QLD 4001, Australia*

Abstract: The use of data assimilation (DA) techniques is well established in environmental sciences, but most applications are non-transferrable and, as such, model-dependent.

We trialled an open source platform to design a DA experiment, tailored to a specific land surface model and geographical context. Once set up, we tested the system by selecting one of a number of generic DA algorithms available.

The specifics of the experiment relate to the Australian Water Resources Assessment Landscape (AWRA-L) model, in which we assimilated lumped, catchment-average satellite-derived soil moisture retrievals for the Campaspe River Catchment, Victoria, in the Murray-Darling Basin, Australia.

We found the use of a generic DA algorithm feasible, although technical challenges may constitute a threshold in the initial design stages of an experimental set-up, and further refinement and analysis of preliminary results is necessary. Further complications with regard to efficiency and speed are anticipated in refining the tested lumped experimental set-up to one with gridded data and multi-variable constraints.

The experiment shows transferrable DA methods offered by an open source platform can be used to set-up a specific model-observation configurations for the Australian Water Resource Assessment system. While this in principle allows for testing of various DA algorithms and configurations to make the best choice, it potentially cuts time and development costs.

Keywords: Data assimilation algorithm, open source platform, AWRA, soil moisture retrieval.

1. INTRODUCTION

In the on-going effort to set-up annual national water accounts for Australia (WIRADA, 2008) a need has been identified to improve on modelled water-related variables by reconciling them dynamically with sequential sets of observations, both at the single point (on-ground) and as spatially distributed fields (e.g. via remote sensing) (Van Dijk and Renzullo, 2011). This practice is commonly referred to as data assimilation (DA). It combines two sources of information, computational models and observations, to exploit their strengths and compensate for their weaknesses. The technique is well established in atmospheric sciences, and increasingly so in land surface modelling/hydrology, evidenced by a wealth of reported DA experiments in the literature over the last decade or so (e.g., Reichle et al., 2008; Crow et al., 2008; Drusch, 2010). Most studies, however, refer to one-off tailor-made solutions for a specific model-observation configuration, which are often difficult to transfer to the next DA experimental set-up.

A recent development in this space is the emergence of open source platforms, which offer an array of model-independent, generic DA software to pick and choose from, allowing the re-use and comparison of different techniques (Weerts *et al.*, 2010). The advantage of saving time and reducing development costs, however, is balanced by the degree of generalization achievable in the DA method applied. While a general notion exists on which DA algorithms work best in which model-observation configuration, the current approach of consensus is primarily one of trial-and-error. In this study, we opted to try a generic DA platform to investigate if such a system can be used to set-up a specific experiment. Part of the research in the Water Information Research and Development Alliance (WIRADA) examines DA techniques and products, which will support and improve Australian Water Resources Assessment (AWRA) system continental water balance models. Here, only the landscape component of the AWRA system, referred to as AWRA-L, is considered.

AWRA-L is a spatially-distributed hydrological model that simulates stores and flows of water in the landscape on a daily time step. Surface observations (e.g., from stream gauges and flux towers) and derived remote sensing products have been used in the model development and tuning (parameter estimation) to ensure that model estimate are to some degree consistent with the "observable" components of the water (Van Dijk, 2010). Nevertheless, like all conceptualizations of physical systems, AWAR-L is prone to modeling errors and biases. These errors are an inevitable consequence of models striving for general applicability across the landscape by achieving a compromise between lack of diagnostic/prognostic capability and over parameterization.

In this study, we aimed to set-up a DA experiment by connecting the AWRA-L model to the OpenDA platform. Subsequently, we tried to assimilate satellite-derived soil moisture retrievals for the Campaspe River Catchment, Victoria, in the Murray-Darling Basin, Australia into the modelled soil moisture forecasts to improve the analysis with respect to the soil moisture retrievals. While the soil moisture retrievals are available as gridded data (0.25 degree), we averaged the gridded data dimension to a lumped catchment-average value, as a first step and for reasons of efficiency/speed of the implemented AWRA-L/OpenDA connection.

2. METHODOLOGY

2.1. OpenDA

OpenDA is a "generic environment for parameter calibration and DA application. It provides a platform that allows an interchange of algorithms and models."(www.openda.org). In a modular framework, it contains methods and tools that can be applied to a wide range of applications, e.g. weather prediction, environmental management, oil exploration, traffic management. It allows re-use of DA software, with the intention of reducing the costs of applying DA methods. By the same token, it facilitates an accelerated uptake of new insights in the field of DA by all applications that might benefit from it.

The OpenDA software has been developed by a cooperation of the Delft University of Technology, research institute Deltares and the scientific engineering company VORtech, all based in the Netherlands. After several years of development and beta testing, the open source version 1.0 of OpenDA was released on 10 May 2010, available through the sourceforge.net website. While unstable elements were purged from the official OpenDa release version, supplementary features and documentation necessary for a broader user community were added. The OpenDA environment has been used in a number of case-studies, e.g. a tide model for Singapore (Kurniawan et al., 2010) and air quality monitoring over the North Sea (Van Velzen and Segers, 2009).

The currently available DA algorithms are the Ensemble method, the Reduced Rank Square method, the Ensemble Square Root method and the COFFEE method, which are all sub-optimal schemes of the Kalman filter. Due to time constraints, we only trialed the first method, as one of the most commonly applied in these types of experiments.

Data assimilation advances by analysis cycles, in which observations of the present state of a system are combined with results from a forecast. This analysis is regarded as the 'best estimate' of the present system state. Next, the model moves forward in time to produce the forecast in the subsequent analysis cycle. While the analysis and forecast may be regarded as probability distributions, Kalman filter equations assume these are Gaussian (normal). The computational requirements for a full Kalman filter are typically prohibitively large, which has given rise to a number of approximations or sub-optimal schemes to reduce processing time. The Ensemble method represents the probability distribution by an ensemble of simulations based on perturbed observations, where the covariance is approximated by the ensemble/sample covariance (Evensen 2007).

2.2. Computational model and state variable

AWRA-L is a physically-based water balance model, that could be regarded a hybrid between a land surface model (LSM) and a lumped catchment model in terms of model structure, complexity and outputs (Van Dijk, 2010). It was developed for the purpose of providing water balance estimates relevant to water resources management at both local and continental scales (Van Dijk and Renzullo, 2011). The model runs at daily time steps and provides gridded estimates at 5-km resolution across the country of the water stores (in the saturated and unsaturated zones) and fluxes (evapotranspiration and runoff). Van Dijk and Warren (2010) conducted a preliminary assessment of an early version of the AWRA-L (v0.5) and found it performed satisfactory. Since then, additional investigations have evaluated the model performance to further develop the AWRA-L modelling capability (Stenson *et al.*2011, Van Dijk *et al.*, 2011, Viney *et al.*, 2011).

AWRA-L does not articulate a physical depth of the top soil layer. This is relevant to our DA experiment, because we assimilate soil moisture retrievals, which relate to a shallow soil depth of ~2 cm. AWRA-L provides top-layer soil water storage (in mm). A maximum soil water holding capacity of 30 mm is used across the Australian continent.

2.3. Observation model and variable

Near-surface soil moisture derived from remotely sensed low-frequency microwave emissions has the ability to improve hydrological and meteorological modelling (e.g. 2002; Koster *et al.*, 2004; Scipal *et al.*, 2005; Crow, 2007). This has led to the development of a number of retrieval algorithms aimed at inferring surface soils moisture from passive microwave brightness temperature (Tb) observations acquired with space borne systems (e.g. Njoku *et al.*, 2003; Owe *et al.*, 2008), of which, until recently, the most widely utilized sensor is the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E).

While passive microwave remote sensing is useful for providing frequent soil moisture data that for most of the globe, the information it contains corresponds to only the top soil layer. Several studies have demonstrated these data can be used to constrain land surface model (LSMs) through data assimilation to provide added value (Walker and Houser, 2001;Koster *et al.*2004; Reichle *et al.* 2008). The satellite data provide a constraint on model estimates of top-layer soil moisture which is subsequently propagated though model simulation to deeper soil layers.

In this study, we use the soil moisture product from the Vrije Universiteit Amsterdam (VUA). The VUA product is derived from low frequency passive microwave satellite observations, such as AMSR-E, with the Land Parameter Retrieval Model (LPRM, Owe *et al.*, 2008) , a radiative transfer-based model. It corresponds to the moisture content in the top ~2 cm of soil.

3. RESULTS

3.1. Coupling of AWRA-L to OpenDA

The AWRA-L model is available in C#/.Net code. At present, OpenDA provides language interfaces for C/C++, Java and Fortran 77/90. As such, the coupling of AWRA-L to OpenDA presents broadly two avenues of approach. First, a platform/OpenDA point-of-view, which requires a re-coding of AWRA-L to one of the compile-able computer languages, C/C++, Java or Fortan 77/90. A second approach converts OpenDA Java

to .Net code, which may be achieved in a number of ways, such as through iKVM, a Java virtual machine application. Alternatively, AWRA-L may be linked through a dedicated C#/.Net interface, which is currently under development at OpenDA. A less attractive option is the coupling of AWRA-L to OpenDA as a black-box model. In this case, the exchange of information takes place via files rather than computer memory (as in compiled code), which significantly slows down the computational process. As such, the black-box coupling is recommended for initial system configuration trialing or parameter calibration rather than operational spatially-distributed state updating.

In this particular experiment, we used a webHttpbinding, which enables the model, here treated as a Web Server exposing AWRA-L, to communicate with OpenDA across HTTP. The webHttpbinding exposes web service functionality allowing interaction between a client (OpenDA) and a server (exposing AWRA-L) via Hypertext Transfer Protocol (HTTP) requests and responses (Microsoft, 2011). AWRA-L is thus accessible to OpenDA via calls to a standard URL (like a web page) and it returns data in the JavaScript Object Notation (JSON) format. JSON is a language independent text based data interchange format, for which .NET provides in-built support through the Windows Communication Framework (Leighton *et al.*, 2011).
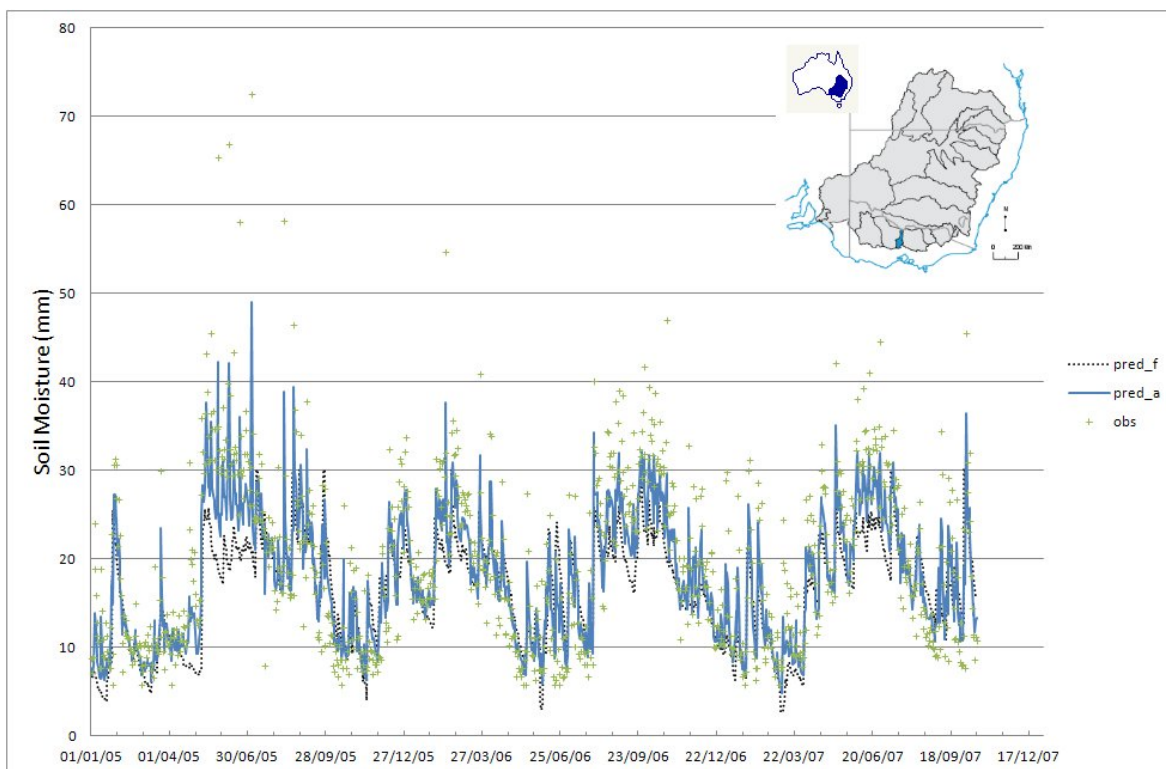

Figure 1 Preliminary results of the assimilation of lumped catchment –average LPRM-derived soil moisture retrievals into the AWRA-L model over a 3-year period (2005-2008) for the Campaspe River Catchment, Victoria (insert) using a 75 member Ensemble Kalman Filter. The analysis (pred_a, blue line) plots in between the retrievals (obs, green cross) and the forecast (pred_f, dotted line).

3.2. Assimilation results

Figure 1 shows results of the assimilation of daily lumped, catchment-average LPRM-based soil moisture retrievals (obs, green cross) into the AWRA-L model forecast (pred_f, dotted line), together with the resulting analysis (pred_a, blue line), using a 75-member Ensemble Kalman Filter over a 3-year period (2005-2008) in the Campaspe River Catchment, Victoria. As a first sanity check, the analysis (pred_a) plots in between the soil moisture retrievals (obs) and the model forecast (pred_f). However, at the peaks the observation far exceeds the model forecast, which pulls the analysis at these peaks at times above the soil moisture threshold of 30 mm in the top soil set in the AWRA-L model. While this may be solved by simply including a threshold in the analysis cycle, it possibly points to overestimation in LPRM-based soil moisture retrievals (Koster *et al.*, 2009). Further, detailed analysis of the interaction between OpenDA and AWRA-L revealed that rather than updating the soil moisture model state, the model output is corrected. This implies the (updated) information is not fed back into the model and, as such, is lost. A further complication arises from the fact the observation (the satellite retrieval) does not directly compare to the model state. In the

lateral sense, this relates to a lumped soil moisture value for the catchment, where the model state is averaged according to the fraction of tall, deep-rooted and shallow-rooted vegetation. In a vertical sense it relates to the shallow reference depth of the soil moisture retrievals (~2 cm) and the model top soil depth, which is, although not specified, certainly deeper. While a gridded data approach reduces the lateral averaging to some extent, the vertical features are implicit to model and observation.

4. DISCUSSION AND CONCLUSIONS

4.1. Discussion

Coupling of AWRA-L to OpenDA

The considerations presented above underscore the preliminary character of the results, which need further work, analysis and refinement. However, it does show HTTP can in principle be used to connect OpenDA to AWRA-L to perform data assimilation, be it in a lumped catchment configuration. In this configuration, a 28 year run for a lumped catchment, consisting of 25 years spin-up and 3 years of data assimilation with a 75 member Ensemble Kalman Filter, takes ~45 minutes on aWindows 7 64 bit, 4GB RAM, CPU:m560 2.67 GHz machine. This may indicate the HTTP connection could prove too slow with the added dimension of gridded data fields and may have to be abandoned to make way for (presumably) faster AWRA-L/OpenDA connections, such as iKVM.

The correction of the model output, as opposed to the updating of the model state, which is subsequently to be fed back in the model, points to a miscommunication between OpenDA and the AWRA-l model. This may indicate some modifications to the AWRA-L code may be necessary for OpenDA, despite its objective of model-independency (Van Velzen and Segers, 2009).

4.2. Conclusions

We set-up a data assimilation (DA) experiment linking the Australian Water Resources Assessment-Landscape (AWRA-L) model to an open source platform. We tested one of a number of available DA algorithms, an (75 member) Ensemble Kalman Filter, to assimilate lumped, catchment-average satellite-derived soil moisture retrievals into modelled forecast for the Campaspe River Catchment, Victoria in the Murray-Darling Basin, Australia. Although the analysis (pred_a) plots in between the soil moisture retrievals (obs) and the model forecast (pred_f), which is an indication of performance, a number of issues remain unresolved, such as the correct identification of the model state in the AWRA-L/OpenDA communication and the accurate pairing of the observation and model state.

The use of generic DA techniques, as offered by open source platforms, is feasible to design a specific, although simplified (lumped), model-observation experiment. Technical challenges, mainly related to linking the model (source code) with the open source platform, however, may constitute a threshold in the initial design stages. Further complications with regard to efficiency and speed may be anticipated when extending the tested lumped set-up to a gridded data experiment and/or multi-variable constraints.

REFERENCES

Crow, W.T, (2007), A novel method for quantifying value in spaceborne soil moisture retrievals, *Journal of Hydrometeorology*, 8(1), 56-67.

Crow, W. T. and R. H. Reichle (2008), Comparison of adaptive filtering techniques for land surface data assimilation, *Water Resources Research*, **44**, W08423, doi:10.1029/2008WR006883.

CSIRO (2008), Water availability in the Murray–Darling Basin, A report to the Australian Government from the CSIRO Murray–Darling Basin Sustainable Yields Project, CSIRO, Canberra.

Drusch, M. (2010), Initializing numerical weather prediction models with satellite-derived surface soil moisture, *J. Geophys. Res. Atm.*, vol. 112, no. 3, p. D03102.

El Serafy, G.Y., H., Gerritsen, S. Hummel, A.H. Weerts, A.E. Mynett and M. Tanaka (2007). Application of data assimilation in portable operational forecasting systems- the DATools assimilation environment. Ocean Dynamics, Vol. 57, 4-5, 485-499 DOI 10.1007/s10236-007-0124-3

Evensen, G., Data Assimilation, the Ensemble Kalman Filter. Springer, 2007

Koster, R.D., Suarez, M.J., Liu, P., Jambor, U., Berg, A.A., Kistler, M., Reichle, R.H., Rodell, M., and Famiglietti, J. (2004), Realistic initialisation of land surface states: Impacts on subseasonal forecast skill, J. Hydrometeorol., 5, 1049-1063.

Koster RD, Guo Z, Yang R, Dirmeyer PA, Mitchell K and Puma MJ (2009), On the nature of soil moisture in land surface models. Journal of Climate, 22, 4322-4335.

Kurniawan, A., Ooi S. K., H. Gerritsen and D.J. Twigt. Calibrating the regional tidal prediction of the Singapore Regional Model using OpenDA. Submitted for publication in Proc. 9th Int. Conf on Hydroinformatics, Tianjin, 7-11 Sept. 2010

Leighton, B.P., Manser, P., Penton, D.J., Shoesmith, J., Stenson, M.P., Vleeshouwer, M.P. (2011), Rapid development of a Source River Web 2.0 interface, MSSANZ/MODSIM Conference Proc., Perth.

Microsoft., (2011) Configuring System-Provided Bindings. In *MSDN Library*. Retrieved July 7, 2011, from http://msdn.microsoft.com/en-us/library/ms731092.aspx

Njoku, E., T. Jackson, V. Lakshmi, T. Chan, and S. V. Nghiem (2003), Soil moisture retrieval from AMSR-E, IEEE Trans. Geosci. Remote Sens., 41, 215– 229.

Owe M, De Jeu R.A.M., Holmes T.R.H. (2008), Multi-sensor historical climatology of satellite-derived global land surface moisture, *J Geophys Res.*, 113. doi:10.1029/2007JF000769

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**, W03423, doi:10.1029/2007WR006357.

Scipal, K., Scheffner. C., and Wagner, W. (2005), Soil moisture-runoff relation at the catchment scale as observed with coarse resolution microwave remote sensing, *HESS*, 9, 173–183.

Stenson, M.P, Fitch, P., Vleeshouwer, J.,, Frost, A.J., Qifeng, B.,, Lerat, J., Leighton, B., Knapp, S., Warren, G., Van Dijk, A.I.J.M., Bacon, D., Pena Arancibia, J.L., Manser, P. and Shoesmith, J. (2011), Operationalising the Australian Water Resources Assessment (AWRA) system, WIRADA Science Symposium Proc., Melbourne, August 1-5.

Van Dijk, A.I.J.M (2010) The Australian Water Resources Assessment System: Technical Report 3. Landscape model (version 0.5) Technical Description, *CSIRO: Water for a Healthy Country National Research Flagship.*

Van Dijk AIJM and Warren GA (2010) Estimation of AWRA-L model land cover parameters from satellite observations. AWRA background paper 2010/3, WIRADA / CSIRO Water for a Healthy Country National Research Flagship.

Van Dijk, A.I.J.M and Renzullo, L.J. (2011), Water resource monitoring systems and the role of satellite observations, HESS, 15, 39-55, doi:10.5194/hess-15-39-2011

Van Dijk A.I.J.M., Bacon, D., Barratt, D., Crosbie, R., Daamen , C., Fitch, P., Frost, A., Guerschman, J.P., Henderson, B., King, E.A., McVicar, T., Renzullo, L.J., Stenson, .M. and Viney, N. (2011), Design and development of the Australian Water Resources System. WIRADA Science Symposium Proc., Melbourne, August 1-5.

Van Velzen, N. and Segers, A. J. (2009), A problem-solving environment for data assimilation in air quality modelling Environmental Modelling and Software - ENVSOFT , vol. 25, no. 3, pp. 277-288, 2010, DOI: 10.1016/j.envsoft.2009.08.008

Viney, N.R., Van Dijk, A.I.J.M. and Vaze, J. (2011), Comparison of models and methods for estimating spatial patterns of streamflow across a spatial domain, WIRADA Science Symposium Proc., Melbourne, August 1-5.

Walker, J.P and Houser, P.R. (2001), A methodology for initializing soil moisture in a global climate Model: Assimilation of near-surface soil moisture observations, *Journal of Geophysical Research*, 106, 11, 761-774.

Water Information Research & Development Alliance (WIRADA), Science Plan, 2 May 2008, http://www.csiro.au/resources/WIRADA-Science-Plan.html

Weerts, Albrecht H., Ghada Y. El Serafy, Stef Hummel, Juzer Dhondia and Herman Gerritsen (2010). Application of generic data assimilation tools (DATools) for flood forecasting purposes. Computers & GeoSciences, Vol. 36, pp 453-463.