

Towards a Daily Soil Moisture Product based on Incomplete Time Series Observations of Two Satellites

H. Jin^a and B. Henderson^a

^aCSIRO Mathematics, Informatics and Statistics, GPO Box 664, Canberra ACT 2601, Australia
Email: Warren.jin@csiro.au

Abstract: There is a strong need for a reliable daily continental soil moisture product that captures water stored in the top soil layer of the land surface and available for evaporation. Soil moisture is an important component of the water balance in hydrological systems. It has been shown that updating soil moisture at daily intervals in a numerical weather model leads to an increase in precipitation forecast skill. Better daily soil moisture has potential for improved hydrological models and improved prediction of water availability, runoff, flood discharge, etc. Furthermore, uncertainty information about soil moisture data will enable hydrologists, ecologists, agriculturalists, and climatologist to model and predict these hydrological and environmental systems more accurately and reliably, especially at large scales.

Soil moisture is however a challenging quantity to be measured. It is difficult to define as it may depend on the context and the observation platform, and is sometimes assessed indirectly as being the difference in balancing other terms like precipitation, runoff and evaporation. *In situ* soil moisture measurements are continuous, accurate and direct but have limited spatial coverage and a high operational cost. They are also difficult to scale up given the substantial local spatial variability. Similarly, airborne observations have limited spatial and temporal coverage, and a high operational cost. Spaceborne satellite observations have wide spatial coverage, are becoming operationally accessible, but are an indirect measure of soil moisture. They are available at different spatial resolutions, and subject to temporal gaps and limitations in some regions. The provision of ongoing reliable daily soil moisture products at continental scale is a challenge task in the face of these features.

There are several methods for characterising uncertainty/error of satellite retrievals in order to evaluate and blend them. Most of these methods rely on validation of the retrievals with *in situ* or airborne observations, which makes it difficult to evaluate satellite soil moisture products at large scales. In addition, uncertainty estimation is complicated by representativeness and scaling errors. These methods normally do not estimate the uncertainty of each data set but are limited to estimating difference between them. For example, the triple collocation methodology has recently been proposed to characterise error structure of remotely sensed observations and to blend them together into a better product. It still cannot handle the temporal gap issue and its performance depends heavily on the selection of a “reference” product.

In order to characterise uncertainty/error structure of satellite data sets, and create an ongoing reliable daily soil moisture product, we focus on a so-called double collocation methodology. It requires only two independent time series of collocated soil moisture observations, although it can be easily extended to handle three or more sources of observations. After exploring temporal structure of *in situ* observations, linear state space models are introduced to estimate the error structure, the underlying soil moisture status, and mapping functions between observations and the underlying status simultaneously. Unlike existing techniques, it does not require the assumption of a reference data set as ‘ground truth’. The temporal gaps are smoothed out via making use of all the satellite observations as well as imposing some appropriate temporal structure in the underlying soil moisture status time series. The proposed methodology is preliminarily demonstrated on soil moisture observations in the Murrumbidgee catchment region, Australia. The observations are retrieved from the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) and the Advanced Synthetic Aperture Radar (ASAR) global monitoring spaceborne sensors. The paper provides discussion on further development and evaluation of this methodology for creating a reliable daily continental soil moisture product.

Keywords: Blending, soil moisture, remote sensing, state space model, error characterisation

1 INTRODUCTION

Soil moisture is an important component of the water balance in hydrological systems. Amongst things it affects the generation of streamflow from precipitation, reservoir management and flood control, irrigation demands, dry land farming decisions, the likelihood of drought conditions, and weather patterns via its influence on the exchange of water and heat energy between the land surface and the atmosphere through evaporation and plant transpiration (Reichle et al., 2007). There is a clear need for a reliable daily soil moisture product in Australia, in part to provide constraints and to update modelled soil moisture within the Australian Water Resources Assessment (AWRA) modelling system that has been developed under the Water Information Research and Development Alliance (WIRADA) between CSIRO and the Bureau of Meteorology, Australia. Soil moisture is however a challenging quantity to measure. Airborne and *in situ* observations are costly and limited in their spatial and temporal coverage. Satellite based observations have broad spatial coverage and are cost-effective, but can be constrained by the size of their footprint and their temporal frequency. These factors make the task of providing a reliable daily continental soil moisture product from existing observations sources a challenging one (Liu et al., 2011).

There are two major directions to create daily soil moisture products for large geographical regions. Soil moisture can be simulated using models, based on an understanding of physical processes and observed meteorology (Reichle et al., 2007; van Dijk, 2010). Data assimilation allows the incorporation of remotely sensed data into model simulations to improve soil moisture estimates, but the final product remains primarily a modelled one. It is unclear whether the final product retains key characteristics (e.g., inter-annual variations) detected by remotely sensed soil moisture (Liu et al., 2011). The second direction, on which the paper focus, seeks to establish a soil moisture product by combining several meaningful sources of remotely sensed observations together. For example, both passive and active microwave satellites and associated retrieval algorithms have been shown to provide useful retrievals of near-surface soil moisture variations. A passive microwave sensor, Advanced Microwave Scanning Radiometer – Earth observing system (AMSR-E) performs better over sparsely vegetated regions, whereas the Advanced Scatterometer (ASCAT) performs better over regions of moderate vegetation density (Liu et al., 2011). Current techniques have successfully illustrated substantial improvements of a blended product over its component products (Das et al., 2011; Liu et al., 2011).

There are still several challenges for blending satellite soil moisture observations in a reliable way. (1) No single satellite has complete coverage over a long time period. All existing satellite retrieved soil moisture products have incomplete time series of observations and contain gaps, e.g., Figure 1(a) where AMSR-E and the Advanced Synthetic Aperture Radar (ASAR hereafter) scan the whole of Australia about every two and four days, respectively. (2) Inherent uncertainty of soil moisture is normally not provided from a satellite retrieval product, nor mentioned in a blended product. Uncertainty information is however crucial for robust decision making and risk management, as well as driving the respective model and data weights in any subsequent data assimilation at a late stage (Crow and van den Berg, 2010). (3) Different products are based on different sensor specifications (e.g., different microwave frequencies) and often represent different aspects or features of the underlying soil moisture status, which prevent direct merging. For instance, AMSR-E and ASAR represent different quantities, namely a volumetric soil moisture and degree of saturation, respectively. (4) Spatial misalignment is common as soil moisture products often have different spatial resolutions or support. The ASAR has a high spatial resolution (0.5km) in comparison to the 25km resolution of AMSR-E. (5) There lack of ‘ground truth’ data for verification, though *in situ* or airborne observations are often regarded as a benchmark (Liu et al., 2011). This work aims at handling the first three challenges.

Most uncertainty/error characterisation methods rely on the validation of the satellite retrievals against *in situ* or airborne observations directly (Draper et al., 2009) or via data assimilation-based evaluation methods (Reichle et al., 2007; Crow and van den Berg, 2010). The latter combines remotely sensed data with a physical model before validating against field observations. It has the potential to evaluate the relative performance of each soil moisture product. Evaluating against field observations are often limited by their spatial coverage, representativeness and possible scaling errors. Evaluating against modelled variables is also possible but it is usually unclear whether the errors derive from satellite observations or the model structure represented. The triple collocation methodology (Scipal et al., 2008) has recently

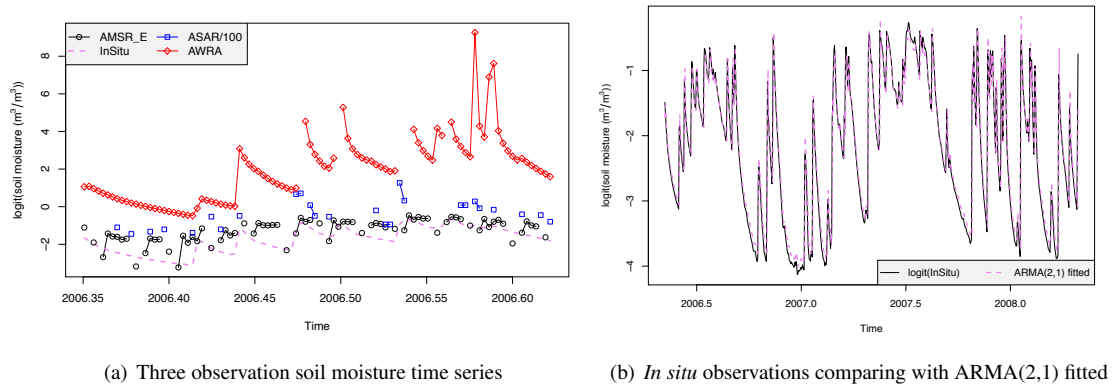


Figure 1. Part of observations from soil moisture Station 7 in Kyeamba region and its related pixel at the logit scale. ASAR data are first divided by 100 to facilitate logit transformation.

been introduced to the field of soil moisture with considerable promise on evaluating remote sensing products without relying on field observations. It allows a pixel-wise estimation of the error structure from the cross-calibration of a set of at least three data sets under the assumption that the data sets are linearly related and that the errors are uncorrelated. The methodology needs to pre-assume a reference data set for tractability, which may or may not be easy to justify. The choice of a different reference data set will lead to different error estimates. None of these techniques abovementioned can handle incomplete time series directly, and they typically require two or more collocated daily observations or modelled data to advance. This is impossible given the latency of the satellites products of interest, e.g., see Figure 1(a).

Considering the incomplete remotely sensed observations, we propose a so-called double collocation methodology. It is based on two (or more) collocated soil moisture time series observations and does not require a lot of co-occurrence at daily time steps. Its feasibility will be illustrated by building up state space models for remotely sensed time series and estimating sensed observation uncertainty, and mapping functions between the observations and underlying soil moisture status via Kalman filtering and maximum likelihood parameter estimation. In addition, a reliable daily soil moisture product can be constructed (Section 3). Some preliminary experimental results will be illustrated (Section 4) for data sets in Section 2. We give concluding discussions in Section 5.

2 DATA SOURCES

Our long term development goal is to provide improved soil moisture input for the Australian Water Resources Assessment (AWRA) system. We will focus on the soil moisture data related with AWRA system. AWRA (van Dijk, 2010) comprises a landscape model, river model and a groundwater model. The most well developed of these is the landscape model. The AWRA landscape model is a spatially-distributed hydrological model that simulates stores and flows of water in the landscape at daily time steps. Surface observations (e.g. from stream gauges and flux towers) and derived remote sensing products have been used in the model development and parameter estimation to ensure that model estimates are to some degree consistent with the ‘observable’ components of the water cycle.

Two remotely sensed soil moisture data sets are considered. The AMSR-E sensor is onboard the National Aeronautics and Space Administration (NASA) Aqua satellite and has provided passive microwave measurements at the 6.9 GHz (C-band) and five higher frequencies since May 2002, with daily ascending (13:30 equatorial local crossing time) and descending (01:30 equatorial local crossing time) overpasses, over a swath width of 1445 km. We use the Land Parameter Retrieval Model (LPRM) algorithm developed by Vrije Universiteit Amsterdam in collaboration with NASA (hereafter VUA-NASA) to retrieve soil moisture as it shows good agreement with *in situ* data over sparsely to moderately vegetated regions (Liu

et al., 2011). The VUA-NASA AMSR-E C-band product, representing the top~1cm of soil, is used. We use the soil moisture retrievals (m^3m^{-3}) acquired by AMSR-E descending passes, as the much smaller temperature gradients between the vegetation and the surface at midnight are known to provide more accurate retrievals. The AMSR-E products are re-sampled, via linear interpolation, into 0.05° (about 5 km) resolution for the period from 1 January 2005 through 31 December 2008.

The second surface soil moisture product, developed by Vienna University of Technology (TU Wien) (Pathe et al., 2009), is estimated from the ASAR global monitoring radar satellite instrument. The ASAR data are at a finer spatial resolution of 0.5km resolution though are only available at a 3-4 day temporal resolution. They represent soil moisture in the top ~1cm of soil. To simplify our study, we upscale ASAR to the same 5km resolution as AMSR-E by taking average value of all the small grid cells in $5km \times 5km$ big grid cells.

Ground based, or *in situ*, observations of soil moisture records spatially distributed real-time measurements of profile soil moisture content. They often serve as bench-marking data to evaluate satellite-borne sensors. We will use the OzNet field data, collected via the Murrumbidgee catchment monitoring network (Young et al., 2008). Continuous near-surface (5cm) soil moisture measurements are available on 38 sites. These observations are averaged out to daily level for the period of 1 Jan 2005 to 31 Dec 2008 to match with the abovementioned data sets. We will focus on the pixels that contain these stations.

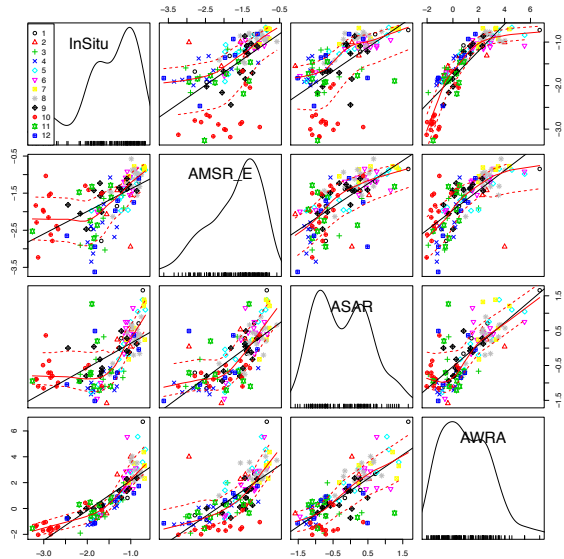


Figure 2. Scatterplot of four collocated soil moisture time series for the $5km \times 5km$ grid containing the Soil Moisture Station 7 in Kyeamba. Only 140 days have data for all the them. Data are logit-transformed and colour coded by month.

3 LINEAR STATE SPACE MODELLING FOR INCOMPLETE TIME SERIES OBSERVATIONS

We implemented the triple collocation methodology at the pixel level soil moisture time series and subsequently sought to relax some of the assumptions it makes, such as the requirement for three collocated products and the specification of a “reference” product. It is however not possible to resolve the error characterisation and blend sensibly due to a parameter identifiability issue when there are only two soil moisture products. We can overcome that by appealing to the fact that we know that the true soil moisture is not fully random and will exhibit some temporal and/or spatial dependence and use that as an alternative and more flexible constraint than the selection of a reference product. Here we are interested in exploiting the temporal dependence or smoothness. The goal of the double collocation methodology is thus as follows: based on two incomplete observation time series to derive, at the same time, an estimate of the underlying true level of soil moisture (i.e., a blended and continuous soil moisture product), the relationship between the true level and each observation product, the uncertainty of each observation soil moisture product.

Time series from the 38 *in situ* soil moisture stations in OzNet were examined and used to identify the most appropriate temporal structure. The longest complete part from the four year window of daily time series was extracted. The data were considered under a logit transformation in order to make data closer to follow a Gaussian distribution. Each time series was then modelled using conventional time series analysis techniques. An Auto-Regression Moving Average (ARMA) was used to identify and diagnose

an appropriate univariate time series model for the *in situ* observation ψ_t at time point t :

$$\psi_t = e_1 \times \psi_{t-1} + \dots + e_p \times \psi_{t-p} + \zeta_t + f_1 \times \zeta_{t-1} + \dots + f_q \times \zeta_{t-q}, \tag{1}$$

where $\zeta_t \stackrel{iid}{\sim} N(0, \sigma_\zeta)$ is a series of normally distributed independent disturbances. In Eq. 1 (ARMA(p, q) hereafter), the non-negative integers p and q denote the number of autoregressive and moving average terms required to model the observed data, and e_i and f_j are coefficients for these terms, respectively.

For most of 38 time series, ARMA(2,1) models are considered to provide reasonable fitting. An ARMA(2,1) model example can be seen in Figure 1(b). This ARMA(2,1) temporal structure is assumed to carry over from *in situ* observations to underlying true soil moisture time series. This does not assume that the *in situ* and the underlying true soil moisture are the same; simply that they exhibit a similar degree of temporal dependence.

We now consider the possible relationship between the remotely sensed observations and the underlying true level of soil moisture. As such a mapping function depends heavily on the soil type, vegetation coverage and meteorological data (like temperature), we consider it for individual pixels separately. Comparing *in situ*, AMSR-E, ASAR, and AWRA modelled soil moisture data on days when all of them exist (see Figure 2 for an example) and taking the *in situ* observations as truth for this purpose, linear or quadratic mappings from the logit-transformed remotely sensed data seem a reasonable assumption.

As the true pixel-wise soil moisture status is not directly observed, one possibility is to model the soil moisture observations via a state space model (Shumway and Stoffer, 2006):

$$\alpha_t = c_t + T_t \alpha_{t-1} + R_t \eta_t \tag{2}$$

$$v_t = d_t + Z_t \times \alpha_t + \epsilon_t \tag{3}$$

where $\eta_t \stackrel{iid}{\sim} N(0, Q_t)$ and $\epsilon_t \stackrel{iid}{\sim} N(0, H_t)$ are white noises; the state vector α_t is driven by deterministic (c_t) and stochastic (η_t) inputs, and its precedent state vector α_{t-1} ; the observation $v_t = [v_{1t}, v_{2t}, \dots, v_{mt}]^T$, the soil moisture observed from m sources at the time t , is linked with the unobserved state vector α_t that includes the underlying soil moisture status ψ_t , with white noise ϵ_t and (possibly) deterministic input d_t . Matrices T_t, R_t, Z_t, Q_t and H_t may depend on a vector of observations v_t , and be time varying or invariant. For simplicity, we will assume these matrices are time invariant and $c_t = 0$. η_t and ϵ_t are assumed serially and also mutually uncorrelated, i.e., $E[\eta_t^T \epsilon_s] = 0$ for all t and s .

We assume linear mapping between the observations and soil moisture status $v_{jt} = d^{(j)} + z^{(j)}\psi_t + \epsilon_{jt}$, with $\epsilon_{jt} \stackrel{iid}{\sim} N(0, \sigma_j^2)$. $\epsilon_t = [\epsilon_{1t}, \dots, \epsilon_{mt}]^T$. We incorporate temporal structure ARMA(p, q) into the state space model. Let $r = \max\{p, q + 1\}$, $e_i = 0$ for $i > r$ and $f_j = 0$ for $j > q$, we have

$$\alpha_t = \begin{bmatrix} \psi_t \\ \sum_{i=2}^r (e_i \psi_{t+1-i} + f_{i-1} \zeta_{t-i}) \\ \vdots \\ e_r \psi_t - 1 + f_{r-1} \zeta_t \end{bmatrix}, T_t = \begin{bmatrix} e_1 & 1 & 0 & \dots & 0 \\ e_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e_{p-1} & 0 & 0 & \dots & 1 \\ e_p & 0 & 0 & \dots & 0 \end{bmatrix}, Z_t = \begin{bmatrix} z^{(1)} & 0 & \dots & 0 \\ z^{(2)} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ z^{(m)} & 0 & \dots & 0 \end{bmatrix},$$

$$R_t = [1, f_1, \dots, f_{r-1}]^T, \eta_t = \text{diag}[\zeta_t, \dots, \zeta_t]_{r \times r}, Q_t = \text{diag}[\sigma_\zeta^2, \dots, \sigma_\zeta^2]_{r \times r}, d_t = [d^{(1)}, d^{(2)}, \dots, d^{(m)}]^T, \text{ and } H_t = \text{diag}[\sigma_1^2, \dots, \sigma_m^2].$$

The model is applicable with two or more observation time series. In this work $m = 2$, i.e., j is either AMSR-E or ASAR.

During the parameter estimation, if either satellite has no observation on a day, we remove its corresponding observation equation in Eq. 3 for that day. If neither of the two satellites has observation, prediction, instead of Kalman filter, is used.

4 SOME PRELIMINARY RESULTS FOR THE MURRUMBIDGEE CATCHMENT REGION

We implement our proposed techniques in R (Shumway and Stoffer, 2006) and derive the maximum likelihood estimate of all the parameters. We give some preliminary results for the pixels that have *in situ*

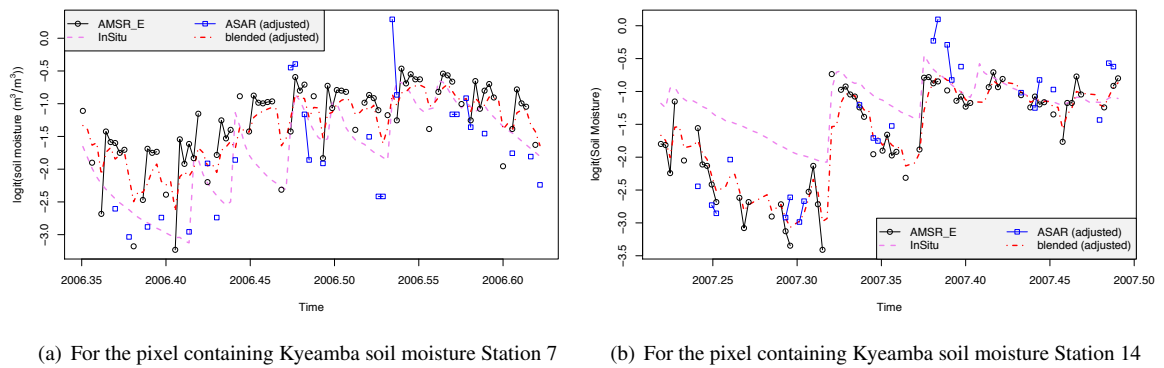


Figure 3. Blended soil moisture time series comparing with three observation ones after adjustment.

soil moisture stations in the Murrumbidgee catchment monitoring network (Young et al., 2008).

The estimated soil moisture status $\hat{\psi}_t$ and ASAR are adjusted via some linear transformations in order to put them on a scale similar to AMSR-E, that is, $\hat{\psi}_t = \hat{\psi}_t \times z^{(\text{AMSR-E})} - d^{(\text{AMSR-E})}$ and $\hat{v}_{\text{ASAR},t} = (v_{\text{ASAR},t} - d^{(\text{ASAR})}) / z^{(\text{ASAR})} \times z^{(\text{AMSR-E})} - d^{(\text{AMSR-E})}$. As illustrated in Figure 3, blended time series are complete at the daily time step, though the two component satellite observation time series are not. A blended time series captures the daily variability in the two observation time series. It normally goes up (or down) when the observations go up (or down). There is alignment with the *in situ* observation time series, though the *in situ* time series does not come into the blending other than providing an estimate of the temporal structure.

In some applications, it is more important that the relative dynamics of soil moisture are reproduced rather than their absolute values (Liu et al., 2011). Thus an appropriate indicator to compare data would be the correlation coefficient instead of the root mean square error (RMSE). For Kyeamba Station 7 (Table 1), the blended time series is highly related with both AMSR-E and ASAR. In addition, it has a slightly higher correlation with *in situ* observation than either of the two remotely sensed time series. It also clearly has a higher correlation with AWRA modelled data than either of the two satellite observation time series. Similar improvement on correlation can be found on Kyeamba Station 14 (Figure 3(b)).

The parameter estimation provides the mapping functions between satellite observations and the underlying status as well as the uncertainty characterisation. Exemplifying on Kyeamba Station 7, $\hat{v}_{\text{AMSR-E},t} = -1.1439 + 0.9327 \times \hat{\psi}_t$ approximately. The uncertainty indicator for AMSR-E $\sigma_{\text{AMSR-E}} = 0.4769$, comparable with $\sigma_{\text{ASAR}} = 0.4795$ for ASAR. For this pixel, both satellite observation time series have similar uncertainty. The uncertainty indicator for the underlying (or blended) soil moisture status $\sigma_{\zeta} = 0.3281$, which is slightly smaller and reflects the benefits of combining the two products. We can also infer the confidence interval for all of these parameters, e.g., the 95% confidence interval is $[0.3529, 0.5686]$ for $\sigma_{\text{AMSR-E}}$. Based on these, we can derive uncertainty information for the blended soil moisture time series.

	blended	AMSR-E	ASAR	InSitu	AWRA
blended	1.0000	0.9204	0.8598	0.7277	0.8394
AMSR-E	0.9204	1.0000	0.6388	0.7230	0.7157
ASAR	0.8598	0.6388	1.0000	0.6694	0.7202
InSitu	0.7277	0.7230	0.6694	1.0000	0.8389
AWRA	0.8394	0.7157	0.7202	0.8389	1.0000

Table 1. Correlation (pairwise) between time series for Station 7 in Kyeamba

5 DISCUSSION AND CONCLUSION

In order to generate a long term reliable daily soil moisture product for Australia, we have proposed a double collocation methodology based on state space models. As demonstrated by our preliminary experimental results, the proposed approach can handle incomplete time series through imposing some temporal structure to the underlying status. This is not a strong assumption. The approach can also estimate the uncertainty information for the component soil moisture observation time series, as well as the complete blended soil moisture time series. Moreover, it estimates the linear relationship between the underlying soil moisture status and each satellite product at each pixel.

We intend to use the double collocation methodology to create a daily soil moisture product following more comprehensive evaluation. We will investigate whether more sophisticated temporal dependency, spatial dependency, nonlinear or non-Gaussian models can further improve the accuracy as they are not uncommon in environmental processes.

Acknowledgement

This work is part of the Water Information Research and Development Alliance (WIRADA) between CSIRO's Water for a Healthy Country Flagship and the Bureau of Meteorology. We appreciate Y.Y. Liu assistance in preparing the soil moisture time series for this work.

REFERENCES

- 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. *Water Resources Research* 46(12), W12519.
- Das, N. N., D. Entekhabi, and E. G. Njoku (2011). An algorithm for merging SMAP radiometer and radar data for high-resolution soil-moisture retrieval. *IEEE Transactions on Geoscience and Remote Sensing* 49(5), 1504–1512.
- Draper, C. S., J. P. Walker, P. J. Steinle, R. A. M. de Jeu, and T. R. H. Holmes (2009). An evaluation of AMSR-E derived soil moisture over australia. *Remote Sensing of Environment* 113(4), 703–710.
- Liu, Y. Y., R. M. Parinussa, W. A. Dorigo, R. A. M. de Jeu, W. Wagner, A. I. J. M. van Dijk, M. F. McCabe, and J. P. Evans (2011). Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydrology and Earth System Sciences* 15(5), 425–436.
- Pathe, C., W. Wagner, D. Sabel, M. Doubkova, and J. B. Basara (2009). Using ENVISAT ASAR global mode data for surface soil moisture retrieval over Oklahoma, USA. *IEEE Transactions on Geoscience and Remote Sensing* 47(2), 468–480.
- Reichle, R., R. Koster, P. Liu, S. Mahanama, E. Njoku, and M. Owe (2007). Comparison and assimilation of global soil moisture retrievals from the advanced microwave scanning radiometer for the earth observing system (AMSR-E) and the scanning multichannel microwave radiometer (SMMR). *Journal of geophysical research* 112(D9), D09108.
- Scipal, K., T. Holmes, R. de Jeu, V. Naeimi, and W. Wagner (2008). A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophysical Research Letters* 35(24), L24403.
- Shumway, R. H. and D. S. Stoffer (2006). *Time series analysis and its applications: with R examples* (2nd ed.). Springer texts in statistics. New York; Berlin: Springer.
- van Dijk, A. I. J. M. (2010). The Australian Water Resources Assessment system. Technical report 3. landscape model (version 0.5) technical description. Technical report, CSIRO: Water for a Healthy Country National Research Flagship.
- Young, R., J. Walker, N. Yeoh, A. Smith, K. Ellett, O. Merlin, and A. Western (2008). Soil moisture and meteorological observations from the murrumbidgee catchment. Technical report, Department of Civil and Environmental Engineering, the University of Melbourne. <http://www.oznet.org.au/>.