

## Variational Gravity Data Assimilation to Improve Soil Moisture Prediction in a Land Surface Model

A. B. Smith<sup>a</sup>, J. P. Walker<sup>b</sup>, A. W. Western<sup>a</sup>

<sup>a</sup>Department of Infrastructure Engineering, Melbourne University, Australia

<sup>b</sup>Department of Civil Engineering, Monash University, Australia

Email: [Adam.Smith@bom.gov.au](mailto:Adam.Smith@bom.gov.au)

**Abstract:** Accurate prediction of soil moisture in a land surface model (LSM) is critical in improving land surface and atmosphere interactions in the atmospheric general circulation models used in numerical weather prediction and global climate models. Gravity is a relatively new source of remotely sensed data, available since the launch of the twin Gravity Recovery And Climate Experiment (GRACE) satellites in March 2002. After correction for other signals, gravity provides a measure of terrestrial water storage (TWS) that includes soil moisture and groundwater. However, assimilating gravity data into a LSM to improve soil moisture prediction has not been widely investigated.

A variational approach to assimilating gravity data is presented for one temperate site in south-east Australia (-35° latitude) for the year 2005. Parameters for the LSM are tuned to the site and forcing is compiled from local atmospheric observations. Gravity observations (corresponding to GRACE products) are synthetically generated from soil moisture and groundwater observations at the site. The model independent parameter optimiser PEST is used to determine initial conditions of soil moisture for a (monthly) moving window by minimising the modelled and observed gravity residual. The assimilated gravity observation is the difference of monthly gravity averages. The ability of variational assimilation to temporally and spatially disaggregate the lumped soil moisture signal present in gravity data is shown.

Variational gravity data assimilation reduces predicted soil moisture (and TWS) bias when compared to the open loop predictions made with degraded LSM forcing data (Table 1). As well as reducing soil moisture bias, gravity data assimilation also improves the predicted soil moisture (and TWS) variance bringing it closer to the observed variance. While variational gravity assimilation retrieves the true (observed) soil moisture variance well, it cannot effectively correct the soil moisture bias in deeper layers resulting from A horizon (average depth of 30 cm) soil parameters being used throughout the profile (to a depth of 4.6 m). It is believed that assimilation performance would improve significantly if more realistic model soil parameters were available.

This paper is a proof of concept, that variational methods can be used to assimilate gravity into a land surface model and it does improve soil moisture predictions.

**Keywords:** Data assimilation, gravity, soil moisture, variational, land surface model, GRACE

**Table 1.** Bias and variance of observed (Obs), open loop (OL), and retrieved (Assim) soil moisture and TWS after gravity data assimilation. Statistics are shown for observed rainfall (595 mm), dry (243 mm) and wet (1219 mm) open loop forcing. RMSE is not shown as it is dominated by the large bias and shows the same trends.

	Bias				Variance			
	0-30 cm	30-60 cm	60-90 cm	TWS	0-30 cm	30-60 cm	60-90 cm	TWS
Obs	0	0	0	0	87.0	20.6	11.1	0.02
OLDry	-9.4	-12.8	-19.8	-0.90	17.5	5.8	0.9	0.00
OLWet	-7.0	-10.7	-18.1	-0.83	23.6	12.1	4.3	0.00
AssimDry	-7.0	-10.5	-17.5	-0.77	33.0	18.2	9.9	0.02
AssimWet	-5.9	-9.7	-17.0	-0.77	30.9	18.9	10.4	0.02

## 1 INTRODUCTION

Accurate prediction of soil moisture by land surface models (LSM) is critical for improved land surface feedback to the atmospheric general circulation models used in numerical weather prediction and global climate modelling. Gravity is a relatively new source of remotely sensed data, available since the launch of the twin Gravity Recovery And Climate Experiment (GRACE) satellites in March 2002. Unlike soil moisture, effective means of assimilating gravity data into a LSM have not been thoroughly investigated.

Recently Zaitchik *et al.* (2008) assimilated GRACE data into a land surface model using a sequential method, the Ensemble Kalman Smoother. A variational approach is presented here, using the model independent parameter estimation software PEST to adjust initial conditions of soil moisture for a moving window by minimising the modelled and observed gravity residual. Variational data assimilation is a natural setting for observations that are temporal averages as this method adjusts all predictions in the assimilation (or averaging) window. The ability of variational gravity assimilation to improve soil moisture prediction is illustrated through a case study. This consists of an open loop model run with degraded atmospheric forcing illustrating a worst case scenario, the success of the soil moisture retrieval is then assessed against observations.

This paper is a proof of concept, that variational methods can be used to assimilate gravity into a land surface model and it does improve soil moisture predictions.

## 2 LAND SURFACE MODEL

The CSIRO Atmosphere Biosphere Land Exchange (CABLE) LSM uses the Clapp and Hornberger (1978) soil water retention curve and corresponding hydraulic conductivity with Campbell's  $b$  soil parameter (Campbell, 1974). The soil water retention curve and hydraulic conductivity are used with the Richard's equation to model soil moisture movement in the profile using six soil layer depths of 2.2, 5.8, 15.4, 40.9, 108.5, and 287.2 cm. Soil temperature is also modelled for each soil layer.

## 3 DATA

The data requirements for soil moisture retrieval from gravity data assimilation into a LSM include:

- atmospheric forcing data to drive the LSM
- data sets to provide soil and vegetation parameters to the LSM
- initial conditions for the LSM (soil moisture and temperature amongst other state variables)
- soil moisture data to assess the soil moisture retrieval (from the gravity data)
- gravity data that is to be assimilated (and that the soil moisture is to be retrieved from), and finally
- perturbed forcing data to create degraded (or open loop) model runs (that will be improved by assimilation).

Further details on the soil moisture and forcing data can be found in Smith *et al.* (2011) and Siriwardena *et al.* (2003) while details on the perturbed forcing (and perturbation method) are given in Pipunic *et al.* (2011) and Turner *et al.* (2008).

### 3.1 Forcing

CABLE requires atmospheric forcing data (that would be provided by the host atmospheric model in a coupled mode) to run offline. This data includes: precipitation, air temperature and pressure, relative humidity, wind speed, and downward short and long wave radiation at hourly (or finer) temporal resolution. The forcing data used for this study is documented in Siriwardena *et al.* (2003) and is primarily data from an automatic weather station (AWS) 20 km north of the study site. The data from January 2000 (earliest available) to December 2004 is used. Local forcing data from an eddy correlation (EC) flux station within 200 m (Pipunic *et al.*, 2011) is used for the year 2005.

Precipitation from the AWS (20 km north) is used for 1 January 2000 to 14 November 2001, a nearby (12 km south) soil moisture site from 14 November 2001 to 5 December 2003, the study site from 5 December 2003 to the end of December 2004, and the flux station (200 m away) for the year 2005. Rainfall is measured at all locations (AWS, two soil moisture sites, and EC flux station) with 0.2 mm resolution tipping bucket raingauges.

### 3.2 Parameters

Global (2° resolution) spatially varying soil and vegetation parameters provided with CABLE are based on the datasets used by Potter *et al.* (1993). Soil parameters are derived from the FAO/UNESCO soil map of the world and given as 9 soil types. Similarly vegetation parameters are provided for 13 vegetation classes. The default soil and vegetation parameters need to be augmented with leaf area index (LAI) which is calculated from remotely sensed (AVHRR) 0.05° monthly average woody and herbaceous fractional cover for the period 1981-1994 (Lu *et al.*, 2003).

The model (soil) parameters that affect soil moisture prediction are tuned for the study site to improve predictive capacity. Soil hydraulic properties are sourced from a high resolution soil survey of the Kyeamba Creek catchment (John Gallant, pers. comm., 2002, CSIRO Land and Water) and are average A horizon parameters for the O'Briens Creek soil unit (ob) as the B horizon parameters are sparse and unreliable (Smith and Zhang, 2007). Parameters sourced from the map include bulk density, saturated hydraulic conductivity, field capacity and wilting point (actually soil moisture at 0.1 bar and 15 bar). Saturation  $\theta_S$  is calculated from bulk density  $\rho_b$  and the equation of Williams *et al.* (1992). Campbell's  $b$  parameter is given by the soil water retention curve after solving simultaneous equations for field capacity and wilting point. Similarly soil suction at saturation is calculated from the soil water retention curve using Campbell's  $b$ , saturation and field capacity.

### 3.3 Initialisation

The LSM (CABLE) is initialised by spinning up ten times over 2000 and run through to the end of 2005. The forcing of Siriwardena *et al.* (2003) is used for 2000 to the end of 2004 and that of Pipunic *et al.* (2011) for 2005.

### 3.4 Evaluation

The LSM (and later the soil moisture retrieval) is assessed against 0-30, 30-60 and 60-90 cm soil moisture data for a valley site (K10) in the temperate Kyeamba creek catchment, part of the Murrumbidgee Catchment in the Murray Darling Basin, Australia (Smith *et al.*, 2011). While a consistent bias is present (more severe in the deeper layers) the model simulates the variability of the soil moisture well, with results similar to open loop (OL) in Figure 1. Furthermore, the observations (with the exception of the wetter portion of the 60-90 cm soil moisture) fall within the model parameter limits of wilting point, field capacity and saturation. The bias in the deeper layers is a result of using A horizon (average depth of 30 cm) soil parameters throughout the soil column to a depth of 4.6 m.

### 3.5 Gravity

Gravity varies both temporally and spatially, and is a function of: latitude and elevation; earth and oceanic tides; subsurface, surface and atmospheric mass distribution and density (Smith *et al.*, 2005). In this study it is assumed that the gravity data has been corrected for all other effects and therefore is a measure of subsurface hydrological mass (or terrestrial water storage) only.

Gravity is synthetically generated from terrestrial water storage (TWS) using the Bouguer slab approximation. This approximation is derived by calculating the gravitational attraction of a vertical cylinder and extending the radius to infinity (Telford *et al.*, 1990). The gravitational attraction of a slab is proportional to its density and thickness. There is a change in gravity if density is held constant with a change in height (useful for representing water table fluctuations), or if thickness is maintained and density varied (useful for modelling vertical soil moisture changes).

If the soil moisture and groundwater are converted to a hydrological mass and summed the resulting

terrestrial water storage (TWS) is related to gravity ( $g$ ) by

$$g = 41.92\text{TWS} \quad (1)$$

where  $g$  is gravity ( $\mu\text{Gal}$ ) and TWS (cm) is assumed to have a constant density ( $1 \text{ g/cm}^3$ ).

Synthetic gravity observations are generated from half-hourly TWS that is averaged over a month. The gravity observations are a difference of monthly averages representing initially expected GRACE observations (Swenson and Wahr, 2002). Note these monthly differences can be computed by post-processing the (monthly) anomaly product actually generated by each GRACE data centre.

As the land surface model (CABLE) does not explicitly model groundwater the predicted gravity (and TWS) is calculated by summing soil moisture over the model depth (4.6 m)

$$g_{\text{mod}} = 41.92 \sum_{i=1}^6 \theta_i z_i \quad (2)$$

where  $\theta_i$  is modelled (volumetric) soil moisture for CABLE layer  $i$  and  $z_i$  is thickness of that layer. The (synthetic) gravity observations are calculated from observed soil moisture and groundwater using a specific yield equal to the maximum observed 60-90 cm soil moisture (an estimate of soil saturation) and the assumption that 60-90 cm soil moisture represents moisture from 60 cm to the water table,

$$g_{\text{obs}} = 41.92 [(\theta_{0-30} + \theta_{30-60}) 30 + \theta_{60-90} (WT - 60) + \theta_{60-90}^{\text{max}} (460 - WT)] \quad (3)$$

where  $\theta_{k-l}$  is observed soil moisture from  $k$  to  $l$  cm and  $WT$  is observed water table depth (positive downwards). Note  $WT$  is always less than 4.6 m.

### 3.6 Perturbed Forcing

Two open loop scenarios are considered. Both use perturbed forcing from Pipunic *et al.* (2011) generated using the Turner *et al.* (2008) method that (stochastically) assigns a prescribed error to each individual forcing variable. Error ranges for each forcing variable based on recommendations of an expert (John Gorman, pers. comm., 2006, Australian Bureau of Meteorology) are used to generate an ensemble of forcing data (Pipunic *et al.*, 2008). From twenty ensemble members two extreme cases are selected, one with 243 mm of annual rainfall (denoted DRY) the other (denoted WET) with 1219 mm, these compare to annual observed rainfall of 595 mm.

## 4 DATA ASSIMILATION

The approach used for the data assimilation is a brute force variational method similar to Sabater *et al.* (2007), using the model independent parameter estimation software PEST (Doherty, 2010). PEST adjusts user specified parameters to minimise the sum of squares of differences between model predictions and observations over a fixed window using the Levenberg-Marquardt algorithm.

As GRACE observations are monthly gravity anomalies (the difference between the monthly observed and climatological average gravity) the window is one month and the observations are differences of average gravity over adjacent monthly windows. The retrieved parameters are initial soil moisture conditions for the window. The window moves through one year (2005) retrieving new soil moisture initial states at the beginning of each month, corresponding to the best fit with observed change in total terrestrial water storage for that month.

The optimisation scheme of the variational assimilation adjusts the (log of) soil moisture for the first time step of the assimilation window. The (initial) soil moisture is adjusted for model layers 1, 4, 5, and 6 only, with the (initial) soil moisture of the second and third model layers computed as a ratio of the (adjusted) soil moisture of the first layer based on that observed at the end (last timestep) of the previous assimilation window. The second and third model layers are tied to the first due to a strong correlation that results in very similar monthly averages for all three layers over the assimilation period.

In addition to ratios of soil moisture (for layers 1 and 2, and layers 1 and 3) from the end of the previous assimilation window, the four soil moisture states (model layers 1, 4, 5, and 6) are also used as a prior

and effectively assimilated as additional observations. The previous soil moisture is used as a prior to make the assimilation smoother, both temporally and vertically throughout the profile. Both using a prior (increasing the number of observations) and tying the top three model layers (reducing the number of estimated model states) improves the identifiability of the states and consequently the performance of the assimilation.

The parameters, a priori initial conditions and forcing of the open loop and assimilation runs are identical. The only difference between open loop and assimilation runs is that gravity observations are assimilated.

## 5 RESULTS

The gravity difference assimilation consistently improves soil moisture over all depths and also TWS estimates (Table 2 and Figure 1). When using gravity differences there are only eleven observations for the year, consequently the open loop and assimilation runs are identical in January. Furthermore, there is virtually no improvement in soil moisture estimates in the drier first half of the year (January to June) due to the negligible amount of month to month TWS change. Modest improvements in soil moisture prediction from the gravity assimilation are evident in the wetter portion of the year (July through to December). While the improvements are larger for the dry open loop case it should be noted that the soil moisture prediction after assimilation is roughly consistent between both the dry and wet open loop cases. Furthermore it is evident that the gravity assimilation retrieves the true (observed) soil moisture variance fairly well, but cannot correct the bias evident in Table 2 and Figure 1. Due to a lack of precipitation the dry open loop variance is underpredicted (compared to the variance of the observations) and this is why the gravity assimilation is more effective in this case. Interestingly, in the wet open loop case the soil moisture is still underpredicted (although the variance is more accurate in the deeper layers) due to A horizon (average depth of 30 cm) soil parameters being used throughout the profile to a depth of 4.6 m. The A horizon soil has a high sand content with a correspondingly high hydraulic conductivity. It is believed significant improvement in gravity data assimilation performance (and open loop model performance) would result from the use of more accurate soil parameters, particularly the hydraulic parameters.

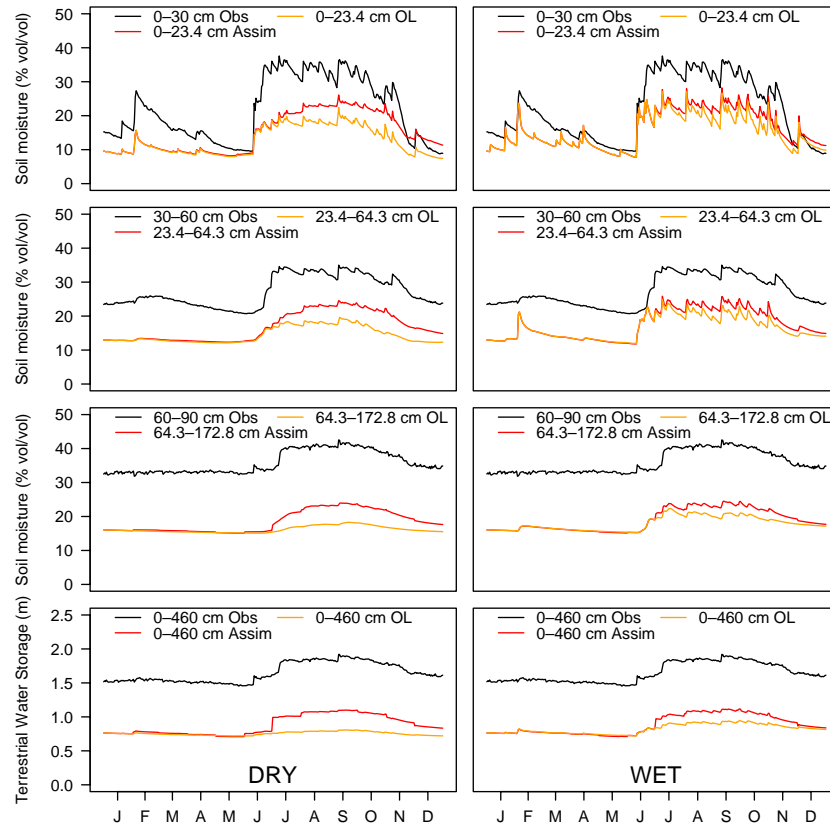
It should be noted that the improvement in soil moisture estimates at the end of the year (last timestep) are larger than at the beginning (due to the assimilation and open loop starting with the same initial conditions), therefore larger improvements in the soil moisture retrieval would be expected in the following (and subsequent years) if the assimilation period were to be extended. The bias evident in open loop predictions is reduced, though still present in the assimilation runs with gravity assimilation clearly unable to significantly improve bias. However, variability of the observations is well retrieved particularly for TWS and 60-90 cm soil moisture (Table 2 and Figure 1).

## 6 CONCLUSIONS

Variational assimilation using gravity difference observations reduced 0-30, 30-60, and 60-90 cm soil moisture and terrestrial water storage (TWS) bias compared to the open loop. Gravity data assimilation also improved 0-30, 30-60, and 60-90 cm soil moisture and TWS variability, bringing it closer to observations than the open loop. It is believed that assimilation performance would improve significantly

**Table 2.** Bias and variance of observed (Obs), open loop (OL), and retrieved (Assim) soil moisture and TWS after gravity data assimilation. Statistics are shown for observed rainfall (595 mm), dry (243 mm) and wet (1219 mm) open loop forcing. RMSE is not shown as it is dominated by the large bias and shows the same trends.

	Bias				Variance			
	0-30 cm	30-60 cm	60-90 cm	TWS	0-30 cm	30-60 cm	60-90 cm	TWS
Obs	0	0	0	0	87.0	20.6	11.1	0.02
OLDry	-9.4	-12.8	-19.8	-0.90	17.5	5.8	0.9	0.00
OLWet	-7.0	-10.7	-18.1	-0.83	23.6	12.1	4.3	0.00
AssimDry	-7.0	-10.5	-17.5	-0.77	33.0	18.2	9.9	0.02
AssimWet	-5.9	-9.7	-17.0	-0.77	30.9	18.9	10.4	0.02



**Figure 1.** Observed, open loop (a model run using degraded forcing), and retrieved soil moisture and TWS after gravity data assimilation. Results are shown for a temperate valley site ( $-35^{\circ}$  latitude) for the year 2005 using a one month assimilation window with dry (and wet) open loop forcing. Assimilation (Assim) and open loop (OL) are in red and orange, observations (Obs) in black.

further with more realistic model soil parameters, as model calibration was not performed in this study.

A gravity difference observation (rather than the absolute gravity value) is used as no knowledge of the period used to generate the background mean for gravity anomalies (available from GRACE data centres) is required. Furthermore, the background mean gravity field used in deriving the products may change in the future as more gravity data becomes available, and LSM forcing data may not be available to generate a modelled background mean for the same period. Consequently, it is recommended that if the period of the background mean is unknown, or data for this period unavailable, gravity anomalies be differenced and the gravity difference assimilated.

The current work could be extended by considering gravity anomalies (rather than differences) such as are currently provided by the GRACE data centres. Furthermore the method should be applied to GRACE data in a spatial setting using a gridded land surface model.

#### ACKNOWLEDGEMENT

This study was funded by Australian Research Council Discovery Grant DP0343778.

#### REFERENCES

- Campbell, G. (1974). A simple method for determining unsaturated conductivity from moisture retention data. *Soil Science* 117, 311–314.
- Clapp, R. and G. Hornberger (1978). Empirical equations for some soil hydraulic properties. *Water*

- Resources Research* 14, 601–604.
- Doherty, J. (2010). *Manual and Addendum for PEST: Model Independent Parameter Estimation*. Brisbane, Australia.
- Lu, H., M. R. Raupach, T. R. McVicar, and D. J. Barrett (2003). Decomposition of vegetation cover into woody and herbaceous components using AVHRR NDVI time series. *Remote Sensing of Environment* 86, 1–18.
- Pipunic, R., J. Walker, and A. Western (2008). Assimilation of remotely sensed data for improved latent and sensible heat flux prediction: A comparative synthetic study. *Remote Sensing of Environment* 112(4), 1295–1305.
- Pipunic, R., J. Walker, and A. Western (2011). Assimilation of multiple remotely sensed data types for improved heat flux estimation: A field demonstration. *Remote Sensing of Environment*. in preparation.
- Potter, C. S., J. T. Randerson, C. B. Field, P. A. Matson, P. M. Vitousek, H. A. Mooney, and S. A. Klooster (1993). Terrestrial ecosystem production: a process model based on global satellite and surface data. *Global Biogeochemical Cycles* 7(4), 811–841.
- Sabater, J. M., L. Jarlan, J.-C. Calvet, F. Bouyssel, and P. D. Rosnay (2007). From near-surface to root-zone soil moisture using different assimilation techniques. *Journal of Hydrometeorology* 8(2), 194–206.
- Siriwardena, L., F. Chiew, H. Richter, and A. Western (2003). Preparation of a climate data set for the Murrumbidgee River catchment. Cooperative Research Centre for Catchment Hydrology Working Document 03/1.
- Smith, A., J. Walker, A. Western, R. Young, K. Ellett, R. Pipunic, R. Grayson, L. Siriwardena, G. Summerell, V. Shoemark, F. Chiew, and H. Richter (2011). The Murrumbidgee soil moisture monitoring network data set. *Water Resources Review*. in revision.
- Smith, A. and H. Zhang (2007). Evaluating CABLE soil moisture predictions in the Murray Darling Basin. In A. Hollis and V. Jemmeson (Eds.), *Physical processes and modelling of the water and carbon cycle: extended abstracts of presentations at the first annual CAWCR Modelling Workshop 27-29 November 2007*, Melbourne, Australia, pp. 15–18. Bureau of Meteorology Research Centre. BMRC Research Report 133.
- Smith, A. B., J. P. Walker, A. W. Western, and K. M. Ellett (2005). Using ground-based gravity measurements to monitor changes in terrestrial water storage. In *29th Hydrology and Water Resources Symposium*, Canberra, Australian Capital Territory, Australia. I.E. Aust. [CD-ROM], 8 pp.
- Swenson, S. and J. Wahr (2002). Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity. *Journal of Geophysical Research* 107(B9).
- Telford, W., L. Geldart, and R. Sheriff (1990). *Applied Geophysics* (2nd ed.). Cambridge: Cambridge University Press.
- Turner, M., J. Walker, and P. Oke (2008). Ensemble member generation for sequential data assimilation. *Remote Sensing of Environment* 112(4), 1421–1433.
- Williams, J., P. Ross, and K. Bristow (1992). Prediction of the Campbell water retention function from texture, structure and organic matter. In M. van Genuchten, F. Leij, and L. Lund (Eds.), *Proceedings of the International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils*, pp. 427–441. University of California, Riverside, CA.
- Zaitchik, B. F., M. Rodell, and R. H. Reichle (2008). Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi river basin. *Journal of Hydrometeorology* 9(3), 535–548.