SMOS brightness temperature assimilation into the Community Land Model

Dominik Rains[1], Xujun Han[2], Hans Lievens[1,3], Carsten Montzka[2], and Niko E.C. Verhoest[1]

[1]Ghent University, Laboratory of Hydrology and Water Management, Ghent, Belgium
[2]Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences: Agrosphere (IBG-3), Jülich, Germany
[3]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA

Correspondence to: Dominik Rains (Dominik.Rains@ugent.be)

Abstract. SMOS (Soil Moisture and Ocean Salinity mission) brightness temperatures at a single incident angle are assimilated into the Community Land Model (CLM)~~, improving soil moisture simulations over the Australian continent~~. Therefore the data assimilation system DasPy is coupled to the Local Ensemble Transform Kalman Filter (LETKF) as well as to the Community Microwave Emission Model (CMEM). Brightness temperature climatologies are precomputed to enable the assimilation of brightness temperature anomalies, making use of 6 years of SMOS data (2010 - 2015). Mean correlation R increases moderately from 0.61 to 0.68 ~~when~~ for upper soil layers if the root-zone is included in the updates. A slightly reduced improvement is achieved ~~when restricting the assimilation~~ if the assimilation is restricted to the upper soil layers. ~~Furthermore, the~~ The long-term assimilation impact is analysed by looking at a set of quantiles computed at each grid cell. Within hydrological monitoring systems ~~,~~ extreme dry or wet conditions are often defined via their relative occurrence, adding great importance to ~~assimilation induced~~ assimilation-induced quantile changes. Although ~~now still limited~~ still being limited now, longer L-band radiometer time series will become available and make model output improved by assimilating such data more usable for extreme event statistics.

1 Introduction

The potential to improve land surface simulations of soil moisture by assimilating information derived from satellite measurements is well known ~~(Mohanty et al., 2017; Chen et al., 2014; Jia et al., 2009; De Lannoy et al., 2007; Parada and Liang, 2004).~~ (Parada and Liang, 2004; De Lannoy et al., 2007; Jia et al., 2009; Chen et al., 2014; Mohanty et al., 2017). Soil moisture products based on data from a number of missions have been used, e.g. ASCAT (Brocca et al., 2010, 2012; Dharssi et al., 2011; Draper et al., 2011) or AMSR-E (Reichle et al., 2007; Yang et al., 2007; Draper et al., 2009) and a combination of both ASCAT and AMSR-E (Renzullo et al., 20 Launched in November 2009, the Soil Moisture and Ocean Salinity (SMOS) spacecraft is the first mission specifically designed to map soil moisture from space (Kerr et al., 2001; Mecklenburg et al., 2016)~~. The on-board~~, the second one being the similar SMAP mission launched in 2015 (Entekhabi et al., 2010b). The passive Imaging Radiometer with Aperture Synthesis (MIRAS) instrument ~~, sensitive at~~ on-board SMOS, sensitive to 1.4 GHz electromagnetic emissions, measures multi-angular top of atmosphere brightness temperatures at horizontal (H) and vertical (V) polarisation~~influenced by, among others, surface soil moisture~~. These brightness temperatures are ingested into a complex retrieval algorithm resulting in soil moisture

estimates (Kerr et al., 2012) readily usable for analysis, input for higher level products or data assimilation. When assimilating these products, which roughly represent the top 5 centimetres of the soil column, into the according model layers ~~(Montzka et al., 2012; Reichle, 2008)~~(Reichle, 2008; Montzka et al., 2012), the assimilation impact in deeper layers will depend on model physics ~~(Montzka et al., 2011; Montaldo et al., 2001)~~(Montaldo et al., 2001; Kumar et al., 2009; Montzka et al., 2011).

5   Alternatively, ~~deeper layers can be updated directly~~ by making use of one of the key advantages of the ~~relationship, i.e. covariance, between observable and unobservable layers by applying methods such as the~~ various implementations of the Kalman Filter (Kalman et al., 1960), deeper and unobserved layers can be updated directly. For plants ~~,~~ these deeper layers act as the root zone, ~~in which~~ where soil moisture has a profound effect on biochemical processes, thus limiting the effect of data assimilation not only to soil moisture (Vereecken et al., 2016). Examples for assimilating SMOS soil moisture retrievals

10 are, among others, given by Martens et al. (2016a), showing that ~~an~~ the GLEAM evapotranspiration model can benefit from assimilating these data over Australia~~and Lievens et al. (2015b), concluding~~, or Lievens et al. (2015b), who conclude that the positive assimilation impact on soil moisture can improve streamflow simulations for the VIC model, as shown ~~for~~ in the Murray-Darling basin. The impact on both streamflow and evaporation is evaluated by Ridler et al. (2014) for western Denmark. ~~More recently,~~ Leroux et al. (2016) assimilate SMOS soil moisture products into the DHSVM model, improving water

15 table depth and streamflow simulations, thereby greatly reducing the uncertainties introduced by the use of uncorrected near real-time precipitation forcings. Scholze et al. (2016) have assimilated SMOS retrievals together with CO2 measurements to constrain the global carbon cycle.

    Apart from assimilating the retrieved soil moisture products, it is also possible to directly assimilate the brightness temper-

20 atures, which should, in theory, eliminate a number of problems. For instance, the SMOS Level 2 ~~and Level 3~~ soil moisture retrievals ~~in essence solve minimisation problems~~ represent the optimum fits between simulated brightness temperatures and the observed satellite signal (Kerr et al., 2012). The simulated top of atmosphere signal ~~is thereby dependent~~ thereby depends on both static and dynamic ancillary data ~~, which is~~ based on input and output of a specific land surface model, e.g. ~~in the case of~~ for SMOS retrievals the European Centre for Medium-Range Weather Forecasts HTESSEL land surface model (Bal-

25 samo et al., 2009). When using a modified or different land surface model it can ~~thus~~ be beneficiary to directly assimilate the brightness temperatures ~~as this allows for a consistent use of~~ in order to use consistent auxiliary information for the land surface model and the radiative transfer model~~(Han et al., 2013). Nevertheless, similar to when~~. In the case of assimilating soil moisture retrievals~~and having to deal with potentially large biases between retrieved and modelled soil moisture, large biases are also common between modelled and observed brightness temperatures due to the many uncertainties involved. The~~

30 ~~source of such uncertainties, among others, may lie in the atmospheric forcings, the land surface representation or the land surface model itself (Drusch et al., 2009; Barella-Ortiz et al., 2015). Since assimilation is expected to correct random errors only and most assimilation algorithms rely on unbiased observations, i.e. bias-blind, it is necessary to remove the bias prior to assimilation (Yilmaz and Crow, 2013). Calibrating the radiative transfer model to closely match the observed time series is one possible solution, as shown by Drusch et al. (2009), De Lannoy et al. (2013) and Lievens et al. (2015a), with the alternative~~

35 ~~being the rescaling of the measurements to mimic more closely the forward simulations (Lievens et al., 2015b). Case studies for~~

assimilation and how to deal with the bias are given by ~~Muñoz-Sabater (2015) and Muñoz-Sabater et al. (2012), both evaluating the assimilation of SMOS brightness temperatures into the ECMWF soil moisture analysis, or by Lievens et al. (2016) who assimilate the soil moisture products as well as brightness temperatures into the Variable Infiltration Capacity (VIC) model for the Murray-Darling basin. Similarly, De Lannoy and Reichle (2016b) compare brightness temperature and soil moisture~~

5   ~~assimilation over the U. S. and De Lannoy and Reichle (2016a) describe the assimilation of only brightness temperatures into the GEOS-5 Catchment Land Surface Model.~~, the auxiliary data used by the model are likely to be correlated with the data used by the retrievals. This inevitably leads to cross-correlated errors between the model and the retrievals, which may have a negative impact on the assimilation performance (De Lannoy and Reichle, 2016a). A selection of brightness temperature assimilation studies is given by Jia et al. (2009); Muñoz-Sabater (2015); De Lannoy and Reichle (2016a); Lievens et al. (2016).

10

    Taken as a whole, ~~studies directly assimilating real brightness temperatures over large scales remain limited and the concept still needs to be further explored, particularly using different qualitative land surface models. Furthermore, the~~ assimilating L-band brightness temperatures in practical terms is quite a new concept which still needs further exploring. The assimilation impact is ~~often evaluated by solely~~ mostly evaluated by comparing soil moisture time series to a limited number of in-situ mea-

15 surements. ~~One study looking more at the long-term assimilation effect, albeit using active microwave data, was carried out by Draper and Reichle (2015), highlighting the fact that despite the required unbiased nature of a bias-blind assimilation system, assimilation can correct for longer-term behaviour and thus be beneficial for monitoring extreme events~~ Given the proven positive impact and the increased availability of longer time series of satellite observations, it is likely that future hydrological monitoring systems, such as ~~droughts.~~ for droughts or floods, will benefit from these data. However, little is still known about

20 long-term assimilation impacts, e.g. on quantiles, which are often used for applications such as drought monitoring.

    Within this study ~~, we specifically look at model state biases that might be introduced into the model over longer time periods, e.g. due to the way model physics potentially react differently to positive or negative increments, model and atmospheric perturbations or unresolved seasonal discrepancies between model and observation. Model state biases~~ can also be more

25 ~~complex in nature than a simple shift of the mean value away from the open loop run and introduce varying changes at different quantile levels within~~ we assimilate SMOS brightness temperatures at H polarisation over Australia from January 2010 until December 2015 into the Community Land Model (version 4.5, Oleson et al. (2013)) and evaluate the assimilation impact both in terms of correlation improvements towards in-situ measurements and in terms of long-term induced model biases, i.e. changes in quantiles, for the state variable soil moisture. The motivation for the quantile analysis is to show how

30 certain behaviour in the cumulative distribution functions (CDFs) ~~. Such changes can be of great importance in the~~ is linked to features in the landscape. We also place the findings within the context of hydrological monitoring systems, ~~since absolute soil moisture values are difficult to compare between grid cells due to e.g. differences in land cover and soil texture. Looking at the relative occurrence of a specific value is more useful, especially when trying to identify spatial patterns, such as areas suffering from extreme conditions like droughts or possible flooding. Examples for existing~~ which mostly use CDFs as a basis

35 to classify areas of interest. A good overview on the evolution of hydrological monitoring systems ~~are for instance the US.~~

~~drought monitor (Svoboda et al., 2002),~~ is given by (Van Dijk and Renzullo, 2011).

We have selected Australia as a study site as we consider it as an ideal test domain for ~~the African Flood and Drought Monitor (Sheffield et al., 2014) or the German Drought Monitor (Samaniego et al., 2013), with all of them using soil moisturequantiles at grid cell level to characterise different levels of severity. With longer~~ long-term brightness temperature assimilation. It is quite heterogeneous in terms of climate and largely uninfluenced by human activity, therefore mostly unaffected by Radio Frequency Interference (Leroux et al., 2013). Although large parts are covered by drylands, the land cover varies along the coastline and includes some densely forested areas in the Australian Alps as well as pasture and areas of intense agricultural activity in the south-east and south-west. The lack of large densely vegetated areas, which mask out the L-band ~~time series becoming available, modelled soil moisture time series improved by assimilation will become sufficiently long for computing the relative occurrence of events and existing monitoring systems, now often relying on purely modelled data , might subsequently benefit from using such data.~~emissions sensitive to soil moisture, is beneficial. Furthermore, soil moisture information based on satellite data is often advertised as being especially useful for monitoring hydrological extremes such as floods and droughts, which Australia is both susceptible to (Johnson et al., 2016; Kiem et al., 2016). A number of L-band specific studies have focused on Australia, covering soil moisture retrieval (Van der Schalie et al., 2015), assimilation studies (Lievens et al., 2015c; Martens et al. validation studies and field campaigns for SMOS (Peischl et al., 2009; Panciera et al., 2008) as well as SMAP (Panciera et al., 2014) and soil moisture downscaling experiments (Piles et al., 2011; Merlin et al., 2012; Dumedah et al., 2015). A comparison of SMOS satellite soil moisture retrievals with products based on other sensors is given by Su et al. (2013). The joint assimilation of passive and active non-L-band microwave data has been tested by Renzullo et al. (2014). More recently, SMOS soil moisture and GRACE water storage have been jointly assimilated by Tian et al. (2017).

~~Within this study we assimilate 6 full years of SMOS brightness temperaturesat H polarisation over Australia from January 2010 until December 2015 into the Community Land Model (version 4.5, Oleson et al. (2013)), and evaluate the assimilation impact both in terms of correlation improvements towards in-situ measurements as well in terms of induced model biases and quantile changes. The CLM is therefore~~ The Community Land Model (CLM) provides all outputs required for the brightness temperature forward simulations, which further motivates the direct assimilation of brightness temperatures. Being part of the fully coupled Community Earth System Model (CESM), it can be used for future coupled land-atmosphere studies using a similar setup as for the brightness temperature assimilation. A full description of the CLM surface data used for modelling the Australian continent is given in section 2.

For the assimilation experiment presented in this study the CLM is coupled to the Community Microwave Emission Model (~~CMEM,~~ version 5.1, Drusch et al. (2009)) forward operator within the data assimilation system DasPy (Han et al., 2015a). The increments are computed with the Local Ensemble Transform Kalman Filter ~~(Han et al., 2015b; Miyoshi and Yamane, 2007; Hunt et al., 20~~ The observation bias ~~problem~~ between forward simulations and observed brightness temperatures is encountered by assimilating anomalies. Remaining differences in mean and variance are resolved by quantile mapping the entire observation anomaly time series towards the ~~offline computed~~ open-loop forward simulation anomalies at each grid point. ~~The assimilation impact~~

4

~~is evaluated by comparing open loop and assimilation results to~~ Details on the implementation of the assimilation system, the forward simulations and the observation treatment are given in section 3.

The in-situ ~~measurements extracted from the International Soil Moisture Network (ISMN) (Dorigo et al., 2011) for two main assimilation experiments: In the first experiment (DA1) brightness temperature assimilation is restricted to the upper three CLM soil layers corresponding to a depth of 9 cm. The upper six model layers, reaching 50 cm, are updated in the second experiment (DA2). These two experiments enable us to examine to what extent CLM model physics are sufficient to propagate upper level increments to the root-zone in comparison to directly applying the increments in this depth.For these two experiments the soil texture perturbations applied for the ensemble generation were incrementally reduced with layer depth, avoiding to large updates in deep layers. Although the analysis is not focused on it, we have included a third experiment (DA0) using homogeneous soil texture perturbations across all layers, highlighting the problem of large increments in lower layers when the ensemble spread is too large~~data used for the validation are from the OzNet and CosmOZ measurement networks (Smith et al., 2012; Hawdon et al., 2014) and were obtained through the International Soil Moisture Network ISMN (Dorigo et al., 2011). For the quantile analysis ~~the~~ quantiles at 1 % steps are computed at each model grid-point, enabling ~~the sufficient~~ a sufficiently precise empirical estimation of the ~~CDFs. As an example to highlight~~ cumulative distribution functions. To exemplify the effects of quantile changes ~~,~~ we show a ~~very~~ dry event defined at the 10 % quantile level and to what extent its ~~classification changes for the open loop run and~~ spatial extent changes when comparing the open-loop run to the data assimilation ~~.~~ results. Part of the experiments is also to show how the CLM translates assimilation updates restricted to the upper soil layers into the root-zone purely through model physics as compared to directly updating both the upper soil layers as well as the root-zone, with the findings being set into relationship to the quantile analysis. The results of the experiments are given in section 4, followed up by the conclusion in section 5.

2   The Community Land Model

The Community Land Model ~~(CLM)~~ is the land surface component of the Community Earth System Model ~~(CESM)~~ and can be run ~~either in coupled mode together with the other CESM components or offline using precomputed~~ offline with pre-computed atmospheric forcings (Oleson et al., 2013). CLM provides global surface datasets which can be interpolated to pre-defined or custom resolutions and grid types both globally as well as regionally, including single point simulations. ~~For this continental scale study , we replace many of the surface datasets with more recent and higher resolution data, creating a consistent surface dataset~~ Interpolating the included surface datasets resulted in artefacts for elevation and grid cell elevation variance as well as Plant Functional Types, with one Plant Functional Type clearly linked to latitudinal borders. We replaced these, but also other surface datasets, with suitable alternatives. For the choice of datasets we kept possible future global applications in mind, which the results of this study could be compared with. At the same time we believe that the Australian continent is well represented by the chosen datasets or that no better suited alternatives were available for the requirements of this study. A description of these datasets is given in the next section. The model resolution was defined at 0.25 ~~degree resolution~~degrees, which agrees

well with the Level 3 observations provided in the EASE 25 km grid. The model is run at 30-minute time steps, with hourly outputs, allowing for ~~the~~ a sufficiently correct temporal alignment of model and satellite observations.

2.1 Surface Datasets

Each grid cell within CLM is divided into land units covering a certain percentage of the total grid cell area. Possible land units consist of vegetation, wetlands, lakes, glaciers and urban areas. Vegetated land units have a single set of soil properties but can be populated by several Plant Functional Types (PFTs), again defined ~~over~~ by their percentage of coverage in respect to the entire grid cell (Bonan et al., 2002). We have updated the model PFTs with information from the Moderate Resolution Imaging Spectroradiometer (MODIS) MCD12Q1 (version 5) land cover products, provided at 500 m resolution in sinusoidal projection and containing a classification of each grid cell describing the dominant plant functional type. On the basis of WorldClim climate data (Hijmans et al., 2005) these plant functional types are reclassified to the ~~CLM compatible~~ CLM-compatible PFTs (Bonan et al., 2002). PFTs were then aggregated to the model resolution, computing the percentage of 500 m pixels of each Plant Functional Type per grid cell. Monthly Leaf Area Index (LAI) values for each PFT within a grid cell were computed by averaging the MODIS 8-daily MCD15A3H (version 6) LAI product, also provided at 500 m resolution in sinusoidal projection, over the assimilation period (2010 - 2015) to derive the monthly climatology and to replace the original climatological LAI values of CLM. The high-resolution LAI values were up-scaled to model resolution by mapping the 500 m pixels to the 500 m reclassified PFT values within each grid cell and subsequently averaging these per PFT. Stem Area Index (SAI) values were also computed on the basis of the high-resolution MODIS LAI data and likewise up-scaled to model resolution, replacing the standard CLM values. Urban and lake areas were extracted from the MODIS land cover information MCD12Q1. Mean topographic height and standard deviation for each grid cell were downscaled from the 3 arc-second HydroSHEDS digital elevation model (Lehner et al., 2008). Soil texture, namely clay and sand fractions as well as organic matter content, were obtained from the global International Soil Reference and Information System (ISRIC) soil database (Hengl et al., 2014) and mapped to the 10 CLM soil layers by nearest-neighbour interpolation according to their respective depths. The ISRIC database provided information on organic matter as the gravimetric percentage of the fine scale soil fraction and we assumed that the coarse scale soil fraction contains no organic matter. Bulk density was used to compute the organic matter content required by CLM, assuming 0.58 g organic matter per kilogram. The rational for creating high-resolution datasets for CLM closely followed the approaches described in detail in Ke et al. (2012) and Han et al. (2012), ~~similarly replacing~~ who similarly replaced the CLM standard datasets.

2.2 ERA-Interim atmospheric forcing

CLM provides forcings (CRUNCEP) ~~not available for~~ which do not cover the required time period. ~~However, the rolling~~ Therefore, due to the release of ERA-Interim reanalysis data (Dee et al., 2011) with a ~~time lag~~ time-lag of only a few months~~enables the assimilation of relatively new satellite measurements, in this case SMOS, and thus ERA-Interim atmospheric~~, these data were used to force the CLM land surface model over Australia.

6

The variables 2 m air temperature, 2 m pressure, short-wave incoming radiation and total precipitation were extracted and specific humidity was computed from the ERA-Interim 2 m dew point temperature and 2 m air temperature. 2 m wind speed was derived from the provided wind speed components in lateral and longitudinal direction. ~~As~~ With ERA-Interim ~~is~~ being produced by assimilating a multitude of observations into an atmospheric model, some of these variables are the result of the analysis step and others of the forecast step~~and~~, thus the data needed to be handled respectively. Forecasts for flux variables are provided bi-daily at 0:00 and 12:00 UTC for 3, 6, 9 and ~~12 hour~~ 12-hour forecast periods and in accumulated form. For example, the precipitation forecast for a 6-hour time window is the accumulated precipitation over 6 hours. In order to obtain a precipitation estimate for the hours 3 - 6, the precipitation forecast for the first 3-hour window needs to be subtracted. This disaggregation was performed for all flux variables to obtain 3-hourly forcing estimates. Analysis variables are valid as instantaneous estimates and no disaggregation had to be performed in their case. The atmospheric forcings were bi-linearly interpolated from 0.75 degrees spatial resolution to 0.25 degrees model resolution. A similar approach for creating atmospheric forcing data based on ERA-Interim, but with additional corrections through ancillary data, is described in Weedon et al. (2014). Time interpolation from 3-hourly to 1-hourly timesteps is performed at CLM runtime ~~applying~~ with an appropriate interpolation algorithm applied to each variable. Incoming radiation is interpolated by using a cosine function simulating the position of the sun~~and~~, for precipitation a nearest neighbour interpolation is used. For the remaining variables linear interpolation is applied.

3 Assimilation system

The assimilation experiments are performed with the open-source multivariate data assimilation system DasPy. Mainly coded in Python, its modular design in principle allows ~~to couple it to~~ the coupling of different models, observation operators and assimilation algorithms. The version used within this study is coupled to the Community Land Model and the Community Microwave Emissions Model (CMEM, de Rosnay et al. (2009)) observation operator. Furthermore, the system uses the Local Ensemble Transform Kalman Filter (LETKF) implementation by Miyoshi and Yamane (2007) for computing the actual increments. Several studies have been performed using DasPy, including the assimilation of synthetic brightness temperatures within the Babaohe River Basin in northwestern China (Han et al., 2012) and in the Rur catchment in Germany (Han et al., 2015c). The system allows for dual state parameter estimation as shown in Han et al. (2014b).

DasPy has been developed with a focus on High-Performance Computing~~and parallelism~~. Parallelism is achieved through ParallelPython, OpenMP, the Message Parsing Interface (MPI) and MPI4Python. Ensemble members can be distributed across different nodes with the core assimilation system, including the LETKF, being confined to one node. Some of the operations are implemented in C++ within the Python environment, using Weave, to further optimise performance. The LETKF itself is a fully parallel Fortran implementation called through F2PY (Fortran2Python).

3.1 Local Ensemble Transform Kalman Filter

The Local Ensemble Transform Kalman Filter (Hunt et al., 2007) is one of the implementations of the Ensemble Kalman square root filter and is deterministic as opposed to stochastic, thus not introducing random noise into the observations. The LETKF

has the advantage over other non-localised implementations ~~,~~ that the analysis performed for each grid point is limited to a local domain, which makes it computationally more efficient and less susceptible to long-range spurious correlations. ~~Although we are using Level 3 data with the antenna pattern already partially accounted for, the~~ The original SMOS footprint is 43 km across and thus covers more than a single model grid cell, which would ~~encourage~~ justify the assimilation in 3D. However, mostly

5   for reasons of simplicity~~and the already~~, and also due to the previously performed inverse distance observation regridding ~~,~~ partially accounting for this, we ~~here only assimilate~~ only use observations directly covering a grid cell.

    Mathematically, the LETKF can be described as follows~~.~~: Model states for each ensemble member k from a total of K ~~ensembles~~ ensemble members are propagated over time by the model M, starting at a previous analysis time step $n-1$, e.g. a previous analysis step within the data assimilation scheme, x^a_{n-1}~~, resulting~~. This results in a new background estimate of the

10   state vector x^b consisting of the soil moisture states for all ensembles at the current time step n.

$$x^b_{n,k} = M_n(x^a_{n-1,k}) \tag{1}$$

The background ensemble perturbations X^b at the current time step can be computed as:

$$X^b = [x^b_1 - \bar{x}^b | ... | x^b_k - \bar{x}^b] \tag{2}$$

The individual ensemble states x^b are ~~propagated~~ mapped into observation space using a forward operator H, in this case

15   CMEM.

$$y^b_k = H(x^b_k) \tag{3}$$

and the forward simulation perturbations are defined as:

$$Y^b = [y^b_1 - \bar{y}^b | ... | y^b_k - \bar{y}^b] \tag{4}$$

Within the ensemble space the analysis error covariance $\tilde{P}^a$ is computed through

20    $$\tilde{P}^a = [(K-1)I + (Y^b)^T R^{-1} Y^b]^{-1} \tag{5}$$

allowing for the computation of $\bar{W}^a$ as the mean weighting vector

$$\bar{w}^a = \tilde{P}_a Y^{bT} R^{-1}(y^0 - \bar{y}^b) \tag{6}$$

resulting in the analysis mean $\bar{x}_a$.

$$\bar{x}_a = \bar{x}_b + X^b \bar{w}^a \tag{7}$$

25   The analysis perturbations are defined as X^a

$$X^a = X^b + W^a \tag{8}$$

with

$$\tilde{W}^a = \sqrt{(K-1)\tilde{P}^a} \tag{9}$$

where the analysis error covariance P^a is given by:

$$P^a = X^b \tilde{P}^a (X^b)^T \tag{10}$$

3.2 Ensemble generation

Model uncertainty is simulated by running the model in ensembles with perturbations applied either to the atmospheric forcings, surface dataset, model parameters or possible combinations of these. In order to account for the model uncertainty in this study, CLM is run with 32 ensembles with spatially-uncorrelated perturbations added to some of the ERA-Interim forcing data, namely air temperature, shortwave radiation and precipitation. Shortwave radiation is perturbed with multiplicative noise with a standard deviation of 0.3, ~~while~~ whereas for temperature additive noise with a standard deviation of 2.5 K is applied. Finally precipitation is perturbed with multiplicative log-normal noise with a standard deviation of ~~0.5. In order to~~ 0.3. The perturbation factors are the same as used by Reichle et al. (2007) and Han et al. (2014a). No spatially correlated noise was added, as the experiments are carried out in 1D, only using one observation per grid cell. To avoid ensemble collapse during dry periods ~~also~~ soil texture is also perturbed once at model startup~~with~~. Here, spatially correlated multiplicative noise ~~and~~ with a standard deviation of 10 percent for clay and sand for the top two soil layers ~~, while constraining the sum of soil and clay to a maximum of 98 percent. For the lower layers~~, is applied. For lower layers the top layer multiplicative factor is rescaled by using the inverse relationship between ~~each soil layers thickness~~ the thickness of each soil layer and the summed soil layer thickness of the two top layers (see Table 1). This ~~should insure~~ is to ensure that increments in lower soil layers do not result in very large changes in soil water in absolute terms, since soil layer thickness greatly increases towards lower layers. ~~Because CLM derives~~ With CLM deriving hydraulic properties based on soil texture, it is to be noted that as a consequence each ensemble member runs with slightly modified model physics. ~~As stated, for demonstration purpose a third experiment (DA0), was performed, using homogeneous texture perturbations for all soil layers~~ Concerning the number of ensembles, an amount of around 30 is common in brightness temperature assimilation studies and should allow sufficient error estimation.

3.3 Observation Operator

Forward simulations from the model space to the observation space are performed with the Community Microwave Emission Model (CMEM, version 5.1). Model output at each observation time, with the observation time rounded to the full hour, serve as input in order to simulate brightness temperatures as measured by the satellite. SMOS ascending and descending orbits have a local overpass time of approximately 6 am and 6 pm. Forward simulations are thus computed at 08:00 UTC on the same day and 20:00 UTC on the previous day for descending and ascending acquisitions respectively, assuming an average time shift of -10 hours for the entire Australian continent. This greatly decreases the number of analysis steps, since individual orbits within one day can be assimilated at once, ~~and it is still assumed to provide~~ assuming that a sufficiently correct temporal alignment

between observations and model forward simulations , is provided. With the western parts of Australia deviating by 2 hours and the ERA-Interim forcings being interpolated from 3-hourly to 1-hourly data we argue that this approach is acceptable.

CMEM requires time invariant information such as soil layer depth, sand, clay and water fractions, surface height as well as the dominant vegetation type covering the grid cell. Plant functional types are reclassified to ECOCLIMAP veg-

5 etation classes and the dominant (Champeaux et al., 2005) and the type for low and high vegetation is then used by the CMEM (Champeaux et al., 2005). Based on this reclassification the LAI information is assigned to the dominant ECOCLIMAP low vegetation classes accordingly. For the offline forward simulations , CLM was run with LAI as daily output in order to make use of the model-internal LAI interpolation, creating a smooth LAI time series based on the monthly surface dataset. This also ensures that the LAI values used for the CMEM forward simulation are the same as those used within CLM during

10 assimilation. LAI values for high vegetation classes are fixed within CMEM and not taken from the CLM input data. Other dynamic fields used in the forward simulations are soil moisture and soil temperature for all defined soil layers and 2 m air temperature. CMEM supports different types of sub-modules for specific calculations. Within this study, the Mironov model (Mironov et al., 2004) is has been chosen for the dielectric constant computation. Vegetation temperature is computed directly by CLM and used as an input without the need of an approximation, e.g. through air temperature. Effective temperature is ob-

15 tained through the Wigneron model (Wigneron et al., 2001) and applied in the dielectric model. For smooth surface emissivity, soil roughness and vegetation opacity, the Fresnel, Choudhury (Choudhury et al., 1979) and Wigneron (Wigneron et al., 2007) models are used respectively. Finally, atmospheric contributions are estimated applying with the Pellarin methodology (Pellarin et al., 2003). For all modules the standard parameters for CMEM 5.1 remained unchanged and the forward observation model was not calibrated. Although the standard parameters are very unlikely to be perfect for the different land cover classes, we

20 argue that this approach is not necessarily worse than the alternative of calibrating the radiative transfer model. By modifying parameters, such as surface roughness, the bias between forward simulations and observations can be removed, but in some cases at the expense of a reduced sensitivity towards soil moisture. Therefore we remove the static bias between simulations and observations through observation rescaling.

3.4 Observations and anomaly computationSMOS preparation

25 Large biases are common between modelled and observed brightness temperatures due to the many uncertainties involved, such as in the atmospheric forcing, the land surface representation, the land surface model itself as well as the radiative transfer model and its parametrization (Drusch et al., 2009; Barella-Ortiz et al., 2015), with this study being no excepetion. The assimilation is expected to correct random errors only, i.e. bias-blind, and it is therefore necessary to remove the bias prior to assimilation (Yilmaz and Crow, 2013). Calibrating the radiative transfer model to closely match the observed time series is

30 a possible solution, as shown by Drusch et al. (2009), De Lannoy et al. (2013) and Lievens et al. (2015a), with the alternative being the rescaling of the measurements to mimic more closely the forward simulations (Lievens et al., 2015b), as mentioned above. The details of preparing the observations prior to assimilation are given here.

SMOS Level 3 daily brightness temperatures at horizontal H polarisation and 42.5 incidence angle provided by Centre Aval de Traitement des Donées (CATDS) are used in the study and processed for the years 2010 - 2015 (version 310). The data

are rigorously filtered by using ancillary data from the corresponding Level 3 soil moisture products (version 300), excluding measurements with a probability of Radio Frequency Interference (RFI) greater than 0.2 and a Data Quality Index (DQX) value greater than 0.07. Measurements with a number of activated science flags, namely strong topography, snow, flooding, urban areas, coastal zone and precipitation, are not considered either. The filtered observation data are regridded from the Equal-Area Scalable Earth Grid 2 (EASE2) 25 km grid to the 0.25 degree rectangular model grid by using inverse-distance interpolation.

On the basis of these data we compute the climatology for each day for the years 2010 - 2015 by averaging along a 7-day moving window across the 6 years, producing separate climatologies for ascending and descending orbits, thus removing seasonal differences between forward simulations and observations (see e.g. De Lannoy and Reichle (2016a)). Anomalies are then computed between the climatologies and the original SMOS time series. Brightness temperature forward simulations based on an open-loop run with 32 ensembles are performed and the ensemble mean climatology is derived in the same way as the observation climatology. SMOS anomalies are then quantile-matched to the ensemble average forward simulation anomalies to account for the differences in variance. The full approach of anomaly computation and quantile matching is to account for seasonal mean differences between simulations and observations and to remove the bias without more aggressive CDF-matching techniques at seasonal level being required. The original brightness temperature simulations over the entire period exhibited a mean warm bias of 21 K for the ascending orbit and 26 K for the descending orbit. Anomaly correlations prior to quantile matching are 0.21 and 0.39 and after quantile matching 0.38 and 0.60 for ascending and descending orbits respectively. Based on the scaling factor between the standard deviation of the original and CDF-matched SMOS anomalies, the observation variance is recomputed. The unscaled observation variance R = 5 K^2 was defined, accounting for a standard instrument error of 3 K and an assumed combined standard mean error of 4 K for the forward simulations and representativeness error. The instrument error can be seen as a low estimate and is based on the assumption that the brightness temperature binning around the 42.5 degree incidence angle results in a slight reduction, when compared to the 4 K instrument error usually applied for Level 1 data.

During assimilation at each time step the current forward simulation is subtracted from the precomputed forward simulation climatology to compute on-the-fly anomalies. The difference between this simulated anomaly and the SMOS anomaly is the innovation, which is used within the LETKF algorithm. The assumption is made that the forward simulation climatology does not significantly change during the assimilation run. In total there are 2063 and 2044 observations for the ascending / descending orbits respectively.

4 Data assimilation and results

In total, three assimilation experiments are carried out

~~are updated making use of the covariances between the ensembles of the topmost CLM layer and the ensembles of all subsequent layers. Within the first experiment (DA1) the top 3 CLM layers, reaching a depth of 9 cm, are updated. The assimilation was not restricted to the top two layers, corresponding to the depth where SMOS is mainly sensitive, since the assimilation impact would have been further diminished, as discussed later on. The top 6 layers, later also referred to as the root zone, are updated within experiment two (DA2). Both experiments are validated~~, updating either top layer soil moisture or both top layer and root-zone soil moisture. Only updating the upper soil moisture allows for testing the ability of the model to feed the assimilation effects into the root-zone through model physics only. Updating the root-zone is carried out with two sets of soil texture perturbations, which largely influence the modelled background error. The objective is to validate the assimilation impact by comparing the ~~assimilation impact to an open loop run in respect to in-situ soil moisture observations. In-situ stations are available for~~ time series before and after assimilation to a number of ~~depths, mostly 5 cm and 8 cm, thus corresponding well to DA1, as well as 30 cm and a limited number of deeper depths, thus corresponding well to DA2~~in-situ measurements. In addition~~to the in-situ validation, quantile shifts in~~, shifts in the soil moisture quantiles in respect to the ~~open loop~~ open-loop run are analysed ~~, highlighting the~~ to highlight some long-term effects of the data assimilation. A set of quantiles is computed at each grid cell ~~allowing for~~ to allow the empirical estimation of the cumulative distribution functions, since ~~varying quantile~~ shifts, both positive or negative, are possible at different quantile levels. ~~For the quantile analysis, a third experiment (DA0) was added using homogeneous soil texture perturbations, as opposed to texture perturbations decreasing~~ Both the impact on correlation and the long-term effects are set into relationship to which layers are updated and to the model background error in the root-zone. The experiments and their results are described in the following and set into context of their potential effect on hydrological monitoring systems, as shown for the exemplary classification of a dry event. We believe this is to be relevant, since L-band data, or data from other sources, are in the long run likely to be incorporated into more and more operational systems.

We would like to note that the spatial patterns of the open-loop soil moisture simulations at different depths were compared to the locally optimised AWRA-L land surface model (http://www.bom.gov.au/water/landscape), concluding that the CLM simulations are plausible.

In the first experiment (DA 1) only the upper three CLM soil layers, corresponding to a depth of 9 cm, are updated. Although the brightness temperatures are only sensitive to soil moisture in up to 5 cm depth, DA 1 was defined as updating the top three layers, since a number of in-situ measurements are taken from a depth of up to 8 cm. For these in-situ sites measurements are also available for deeper layers and we thus define top layer soil moisture as the soil moisture updated in DA 1. The upper six model layers, reaching 50 cm soil depth, are updated in the second experiment (DA 2). We refer to the lower three of these soil layers as the root-zone. These two experiments enable us to examine to what extent CLM model physics alone are sufficient to update the root-zone through the effects of the assimilation on the upper layers, as in comparison to directly applying the increments in this depth. For the experiments DA 1 and DA 2, soil texture perturbations were incrementally reduced with layer depth, ~~showing how sensitive the quantile analysis and the entire assimilation is from these perturbations . Finally, the implications of quantile changes were showcased for one specific dry event and placed within the context of hydrological monitoring systems.~~ minimising the impact of potentially large updates in deep layers. Since increments are computed in

relative soil moisture, identical increments affect absolute soil water very differently, greatly exaggerating the assimilation impact for deeper layers. Perturbations for the two top layers remain unchanged, thus not decreasing the ensemble spread for the layers where SMOS is sensitive to soil moisture. The soil texture of the subsequent layers is perturbed by decreasing the perturbation factor by the inverse ratio between the respective layer thickness and the layer thickness of the two top layers (see ensemble generation under section 3 and Table 1). Within a third experiment (DA 0), homogeneous soil texture perturbations are applied across all layers, highlighting the problem of large increments in lower layers. As will be shown, the larger ensemble spread in DA 0 actually further improves the correlation with in-situ measurements, but at the expense of introducing strong long-term effects. For all experiments the brightness temperature forward simulations are computed by using the CLM output of all layers. The L-band simulations are thereby mostly affected by the output of up to 5 cm depth, which corresponds to the sensitivity of the SMOS sensor.

4.1 Correlation with in-situ observations

For validation, hourly CLM soil moisture output is compared to in-situ measurements obtained from the International Soil Moisture Network ISMN (Dorigo et al., 2011). The original sources of the in-situ data over Australia are the OzNeT (Smith et al., 2012) and CosmOZ (Hawdon et al., 2014) soil moisture observation networks. OzNet in-situ measurement probes are located within the Murrumbidgee catchment in south-east Australia, a limited spatial domain which does however cover a range of different land cover classes representative for Australia (Smith et al., 2012). The Murrumbidgee catchment was also chosen as a site for a SMOS validation campaign (Peischl et al., 2012). Measurements within the OzNeT network are taken with TDR-probes at shallow levels, mostly 5 cm or 8 cm, and at deeper layers, mostly 30 or 60 cm for the sites used in this study. The in-situ measurements that are part of the CosmOZ network are taken using cosmic-ray neutron probes and are therefore representative for a larger horizontal footprint than the more traditional measurements. CosmOZ measurement sites are located within the Murrumbidgee catchment as well as at selected locations close to the Australian coast. Besides the original description of the measurement networks, Renzullo et al. (2014) offer a good overview of the measurement sites within the frame of a soil moisture data assimilation study. In the current study, we did not differentiate between types of measurements due to the small amount of measurements available. For each in-situ measurement the weighted average of the corresponding CLM soil moisture layers is taken, with the layer thickness being used as the respective weights. Figure 1 shows the location of the Murrumbidgee catchment.

Taking into account only measurements with at least one year of data, not necessarily consecutive, correlations improve from 0.613 for the open-loop run to 0.640, 0.678 and 0.681 for DA 1, DA 2 and DA 0 for top layer soil moisture (number of stations, n = 17). Root-zone soil moisture improvements are smaller, with average correlation coefficients of 0.626, 0.644 and 0.648 for DA 1, DA 2 and DA 0 compared to 0.601 for the open-loop run (n = 31). For the upper level soil moisture, correlation improves for all in-situ measurement stations, whereas for the root-zone soil moisture a single in-situ station shows a deterioration of

correlation for DA 1. In the case of DA 2 and DA 0 the correlation at two stations, albeit at different ones, slightly deteriorates. On average, upper soil moisture behaviour thus improves by ~~also~~ additionally updating deeper layers~~and~~, whereas deeper layer soil moisture is slightly enhanced through only ~~assimilating~~ updating top level soil moisture, with the assimilation effects only being ~~propagated~~ applied through model physics. ~~Overall~~All in all, updating the top ~~6~~ six CLM layers results in the largest improvements~~. Figure 3 shows~~, even more so if the identical soil texture perturbations are applied to all soil layers within experiment DA 0, thereby increasing the assimilation impact through an increased ensemble spread and background error. The individual in-situ measurements around the area of the Murrumbidgee catchment and the respective correlation changes for all three experiments towards the open-loop run are shown in Figure 2. For top layer soil moisture the largest improvements are visible for the sites located in the centre of the catchment (OzNet Yanco site) with clear improvements for DA 2 and DA 0 when compared to DA 1. In the case of the root-zone, multiple measurements at different depths were averaged using the CLM layer thickness as weights. Here improvements are also highest for the Yanco site, except for one measurement location showing a deterioration of correlation for DA 0.

Figure 3 shows The Taylor diagrams for the ~~validation results of experiment DA2~~in-situ validation of experiment DA 2. As opposed to Figure 2, all original measurements are included with no vertical aggregation performed. The Taylor diagrams ~~further show~~reveal a slightly decreased normalised standard deviation ~~of the assimilation results in respect to the in-situ measurements~~ when compared to the ~~open loop time series~~~~in respect to the in-situ data~~open-loop time series. In terms of standardised RMSE it is less conclusive, ~~and RMSE is slightly reduced by assimilation~~ with RMSE being slightly reduced for some stations and slightly increased for ~~some others~~others, but never significantly. These findings correspond well to ~~experiment DA~~ experiments DA 0 and DA 1 ~~, not shown~~here(not shown).

The results motivate us to believe that the assimilation system has been sufficiently well designed to improve soil moisture behaviour across the Australian continent, both for top layer soil moisture and the root-zone. However, as with most assimilation studies, validation sites are sparse and do not cover the many-fold combinations of soil texture, land cover, climate etc. which might all have an impact on the assimilation performance. The representativeness error of the in-situ measurement equally remains a problem, with the spatial support of the measurements, in the case of TDR probes only point measurements, being smaller than the area covered by satellite. The assimilation system is therefore not designed to remove the relative bias between simulations and observations, since the exact truth remains unknown, but to improve the temporal behaviour of the simulations, which has been achieved here.

4.2 Soil moisture increments

~~In a bias-blind assimilation system it can be expected that, especially over longer time periods ,~~Especially over long time periods the mean increments ~~should be~~ in a bias-blind assimilation system can be expected to be very close to zero. Figure 4 and Figure 6 show the mean soil moisture increments over the assimilation period 2010 - 2015 for the experiments DA 0, DA 1 and DA 2, separately for the ascending and descending orbits. Both the top three soil layers and the root-zone soil layers were averaged. Distinctive areas of mean positive increments for the ascending orbit are ~~located in the North, South-West and South-East~~ visible in the north, south-west and south-east of Australia, seemingly being linked to the occur-

rence of vegetation ~~. Nevertheless, the positive bias does not~~ (see Figure 1). The areas in the south-west and south-east as well as in the north correspond well to the subtropical, temperature and tropical climate zones respectively and are subject to higher precipitation than the dry areas inland, although other areas along the coast have similar precipitation. The areas in the south-west and south-east correspond to the wheat growing areas of Australia. The misrepresentation of these areas through the CLM surface datasets, such as the use of climatological LAI instead of actual LAI, might well be the source of such patterns. Also, satellite-based estimates of Leaf Area Index are not error-free, as has been shown specifically for the Murrumbidgee catchment (McColl et al., 2011). Irrigation, which is predominantly applied within the south-east, could be an additional source of error, since the forward simulations will be based on seasonally too low soil moisture, causing an incorrect estimation of the brightness temperature seasonality and the subsequent anomaly computation.

Nevertheless, for the ascending overpasses, the positive biases hardly exceed 0.5 % soil moisture and the remaining parts of Australia either show no mean increments bias or slightly negative values~~. These~~, both in areas covered by mostly sparse vegetation and the inner drylands. For the descending orbits the patterns are still visible ~~for the descending orbit or the~~ but are weaker, both for top layers and the root-zone~~soil moisture but are on overall weaker, except for~~. The exception is DA 0~~with~~, where little differences between top layer and deep layer increments are noticeable. Interesting to highlight ~~,~~ is the fact that top layer deviations from zero are strongest for the assimilation experiment ~~DA1~~ DA 1, compared to DA 0 and ~~DA2, with the latter two updating both~~ DA 2, which both update top layer soil as well as ~~the root-zone. A possible interpretation might be that the assimilation into~~ root zone. The reason may be that updating deeper layers results in a ~~longer~~ more lasting effect, ~~potentially~~ moving the model closer to subsequent observations and thus reducing subsequent increments.

Figure 5 and 7 show the increment standard deviations~~, which are substantially stronger for the ascending orbit. Top~~. Spatial patterns of the assimilation impact are very distinctive but do not necessarily correspond to the patterns seen in the mean increments, although they do partly match, as for the south-west and south-east. The relatively large increments in the western Wheatbelt and the Murray-Darling basin, some areas close to the western coast of Queensland and the eastern coast of the Northern Territory show a standard deviation of 2.5 %. The areas in the north seem to be consistent with the occurrence of tussock grasses, as shown by the Australian National land cover map (http://www.ga.gov.au/scientific-topics/earth-obs/). Minimal or zero increments in all layers, especially along the eastern coast, are due to a lack of observations, as these were removed due to the active vegetation science flags, or the fact that the high LAI values for high vegetation prescribed within the forward operator mask all signals from soil moisture.

Concerning the different assimilation experiments, top layer increments are ~~larger~~ largest for DA 1, followed by DA 2 and then DA 0. ~~Deep layer increments are much larger for DA 0 when compared to DA 2. For~~ Being the most dominant dynamic factor for the ensemble generation, precipitation ~~is the most dominant dynamic factor and leads to immediate increased ensemble spread , andthus a~~ leads to an immediate increase in ensemble spread and, as a consequence, to a larger background error ~~,~~ for the very shallow soil layers, thereby also increasing the observation impact. This impact on the ensemble spread will however be dampened and temporally lagged for deeper layers. ~~The static soil texture perturbations, although their effect are also dynamically dependent on the current soil moisture state, lead to model uncertainty not directly induced by the forcing perturbations. Soil texture perturbations decrease with layer depth for DA 1 and DA 2 (see perturbation scaling shown in Table~~

~~1) and lead to the reduced covariance between top layer and deep layers, which results in decreasing increments for lower~~ ~~layers. It is thereby important to note that in terms of absolute soil moisture, large increments in low layers remove or add~~ ~~vastly larger amounts of water than similar increments in shallow layers and will exhibit strong effects on all above layers~~ ~~(see soil thickness in Table 1). As an example, removal of water in a deep lying layer will lead to increased percolation in all~~ ~~above lying layers, resulting in a sudden drying out of the top soil layers. Small biases inherent in the data assimilation system~~ ~~thus might be largely exaggerated when allowing to large deep layer updates.~~ With increasing layer depth, the soil texture perturbations play a more important role in determining the background error. This is visible for DA 0 where homogeneous soil texture perturbations were applied across all layers and increments in the root-zone are not significantly ~~reduced for the root-zone~~ smaller than for the layers above (Figure 5). As a contrast, root-zone increments applied within experiment DA 2 are far smaller than for the upper layers. For the validation with in-situ measurements we showed that these larger increments for DA 0 actually result in a slightly increased correlation over DA 1.

Concluding on the ~~relationship between~~ behaviour of increment bias and increment standard deviation, it seems that there is ~~either a decreased increment bias at the expense of larger deep layer increments, or vice versa. Finally, regarding the spatial~~ ~~patterns of the increments, areas with very low increments in all layers, especially along the east coast, are either based on~~ ~~a lack of observations, as these were filtered due to the active vegetation science flags, or the fact that the high LAI values~~ ~~for high vegetation prescribed within the forward operator mask any signal from soil moisture.~~ a relationship to root-zone updates. Both increment bias and increment standard deviation are largest for DA 1, where the root-zone is not updated at all. The top layer increment standard deviation decreases for DA 2, whilst also updating the root-zone, with a slight decrease of the increment bias. Compared to DA 2, the increment standard deviation is larger in the root-zone for DA 0 and the top-layer increment bias decreases substantially.

As for the differences between the ascending and descending orbits, we conclude that they can be partly explained by referring back to the fact that soil moisture retrievals are of higher quality for the ascending orbits (Hornbuckle and England, 2005; Kerr et al The thermal equilibrium within the soils, which is more pronounced at 6 a.m. local time, reduces the error in the forward simulations and thus increases the covariance between observations and forward simulations. The mean bias between the forward simulations and observations is $5\,K$ less for the ascending orbits, which supports this explanation.

4.3 Soil moisture quantiles

~~In addition to~~ Apart from looking at the increments, we compute a set of quantiles at 1 % intervals for each CLM soil layer and each grid point, both for the assimilation experiments and the ~~open loop~~ open-loop run. Although in principle the assimilation system should be designed bias-free with similar positive and negative increments, the previous section ~~showed~~ has revealed that small increment biases do exist, potentially ~~resulting in long term~~ causing long-term effects in the resulting analysis. Figures 8, 9 and 10 show the 10 % quantile changes, thus very dry conditions, in relation to the open-loop run for the top ~~9 CLM layersand to which extent it shifts for each grid cell when compared to the open loop run~~nine CLM layers. For experiment DA 1 (Figure 9), assimilation has ~~a small~~ an impact on the topmost soil layer with the quantile increasing by ~~up to 5~~ by a maximum of ca. 1 % for large areas and by up to 4 % for spatially very limited areas. Much smaller changes are

visible for ~~deeper layers but with changes nevertheless reaching layers that are not directly updated.~~ the second and third layer, with some areas also showing a negative impact by up to 2 %. CLM model physics result in changes being also visible within the root-zone, layers three to six, and below. One of the visible patterns is again south-east Australia. For the very deep layers some independent patterns emerge which are not visible for the above layers. Most notably in the Nullarbor plain, on the coast

5   of South Australia, the 10 % quantile increases by up to 2 %. Such patterns are related to strong singular increments in very dry areas which accumulate in the deepest layers. Due to the low temporal dynamics in these lower layers, any added water will have a lasting effect especially on lower quantiles. For experiment DA 2 (Figure 10), ~~whilst also updating layers 3 – 6, slightly~~ with the root-zone also being updated, larger impacts on the quantiles in deeper layers can be observed~~with more areas showing an increased 10 % quantile~~. For the ~~largest part , the patterns , at least for upper layers,~~ most part the patterns reflect

10   well the ~~patterns identified in Figure 6, showing the mean increments, although in deeper layers also some independent patterns emerge that cannot be directly explained by any increments bias. Figure~~ ones identified for DA 1. Figure 8 shows the impact on the 10 % quantile for ~~the~~ experiment DA 0 ~~, using~~ with homogeneous soil texture perturbations ~~across all layers. Here, much larger~~ being used. As expected, significant effects are visible especially within the root-zone~~soil, increasing the quantiles~~ , with quantiles being decreased over wide areas of the Australian inland by up to 5 %. ~~Hard to pinpoint the exact mechanisms for this~~

15   ~~behaviour, it does highlight the potential implications of updating thick soil layers.~~ This is the result of the mean increments being slightly negative for inland Australia, which has a large effect when allowing large updates. Also, since absolute soil moisture increases with layer depth due to increasing layer thickness, removal of water in low layers increases drainage in the above layers, resulting in these to dry out. This is especially visible for layers two and three, where inland Australia to a far greater extent shows a lowered 10 % quantile in comparison to DA 1 and DA 2. For the lowest layer a clear positive quantile

20   shift is visible in the area of Lake Eyre, again showing that some patterns are clearly linked to features in the landscape. The land cover map in Figure 1 shows this as the only area that is classified as bare soil although it is mostly a salt plain with water levels of the lake itself being strongly seasonal and the strong increments might be due to flooding, although this was not closer examined.

   Figures 9 - 10 focus on the changes of the 10 % quantile. However, the spatial patterns identified do not necessarily reflect

25   changes at other quantile levels. The complex nature of these shifts ~~at~~ throughout the entire CDFs is shown in Figure 11. The continental average empirical cumulative distribution functions are plotted for ~~the~~ soil layers 1 - 6 for the ~~open loop~~ open-loop run as well as ~~DA1 and DA2. As shown for the 10 % quantile, lower~~ for DA 1 and DA 2. Lower quantiles are increased on average through data assimilation ~~, whereas a small decrease in the upper quantile values can be observed resulting in an overall~~ although at extremely low levels the behaviour tends to reverses again. Here the quantiles decrease when compared

30   to the open-loop. For the upper quantiles a small decrease can also be observed. The point where the decrease switches to an increase, and the data assimilation on average has a neutral impact, is roughly the 50 % quantile. For the subsequent layers this point decreases towards the 40 % quantile. Although DA 0 resulted in the best correlation with the in-situ measurements, it was disregarded at this point since the assimilation impact was too disruptive by strongly drying out the model across many layers.

For DA 1 and DA 2, the interplay of the quantile changes at the various levels results in an average decrease of the standard deviation of the soil moisture analysis, which to a certain extent could be attributed to the anomaly rescaling. First of all, due to the low sample count any quantile-mapping procedure is challenging around the extremes of the distributions. Additionally, the exact observation error for each observation is unknown and although expected to be zero on average, with a small sample size the observation error might on average deviate from zero, affecting the rescaling, since the true limits of the observation anomaly CDFs are unknown. However, many other reasons will equally play a role and as it has been shown, increment bias and standard deviation is linked to certain geographic areas. The disentanglement of the desired systematic enhancements from erroneously introduced effects remains a challenge. There is still no perfect approach to rescaling the observations to match the model or to calibrating the forward observation model. Looking at long-term CDF changes induced by assimilation can be part of evaluating these different approaches with the final application of analysis data to be kept in mind.

Finally, as an example we want to place quantile behaviour within the context of possible hydrological monitoring systems which directly make use of grid cell quantiles and empirical CDFs. The correction of short-term behaviour alone, i.e. hourly or daily, has a minor effect when analysing phenomena that spread across larger spatial scales and time intervals, although large increments, e.g. due to corrected precipitation during a storm, can have an effect on the start and end point of an observed phenomenon, such as a drought. When classifying such an event defined at a specific quantile level, there will be a twofold impact from the assimilation: The changes in the quantile of interest as well as the change in soil moisture itself. Here we highlight a sample dry event on the east coast to show to what extent its classification changes through the assimilation impact. Figure 12 therefore shows root-zone soil moisture at or below the 10 % quantile level for the open-loop run as well as the data assimilation experiment DA 2 for soil moisture conditions in early 2010, thus at the beginning of the assimilation period. Due to the higher 10 % quantile for DA 2, as seen in Figure 10 and 11, the spatial extent of the cluster for DA 2 is reduced but the spatial patterns of soil moisture remain largely the same. At some time periods, not shown here, a higher degree of noise is noticeable within the assimilation dataset. This is likely due to the fact that non-spatially correlated noise was applied to the meteorological forcings, resulting in a heterogeneous background error field for grid points. We thus conclude that despite having carried out the assimilation in 1D, spatially correlated noise is recommended for such applications. An alternative would be to further increase the ensemble size, but at the expense of higher computational resources. Additionally, when trying to extract meaningful statistics on the occurrence of events, such

as droughts, by extracting these as clusters of grid cells over the spatial and temporal domain, it might be ~~especially~~ particularly important to clean up the dataset in the case of data assimilation, using simple filter algorithms, such as applied by Herrera-Estrada et al. (2017). We want to highlight the fact that the shown event is for demonstration purpose and not linked to any major drought.

5 Conclusion

The Community Land Model was set up for the Australian continent and ~~we~~ coupled to the Community Microwave Emission Model. We have substituted the surface datasets with higher resolution and more recent data. ~~Further,~~ Additionally we have replaced the offline ~~forcing data with~~ forcings with the ERA-Interim reanalysis~~data~~. The assimilation over 6 full years, from 2010 – 2015, of SMOS brightness ~~temperatures~~ temperature anomalies with the LETKF improved soil moisture simulations when compared to in-situ measurements in the order of maximum 11 % ~~, which~~ for top soil moisture. This is similar to the impact observed in other studies. ~~The CLM model remained uncalibrated, as this is tremendous task for large areas and calibration towards soil moisture simulations is difficult for the lack of in-situ measurements , but given the results, we are confident that this specific CLM setup could be applied to larger areas or at global scale. CLM model physics alone did propagate assimilation effects into the root zone when restricting assimilation to the top three layers, improving the correlation with in-situ measurements both for the root zone and surface soil moisture. However, the improvements were largest when directly updating both surface soil moisture and root zone soil moisture~~Both the CLM model and the forward observation model remained uncalibrated.

In detail, three data assimilation experiments were carried out: Within the first experiment the top two layers were updated which correspond to the layer depth were SMOS is sensitive to soil moisture changes. The correlation with top layer soil moisture measurements increased by ca. 5 %, root-zone soil moisture by 4 %. Within the second experiment both top soil moisture and root-zone were updated, resulting in correlation improvements of 11 % and 7 % respectively. The CLM is therefore able to translate top layer updates into deeper layer soils, while restricting the updates to the top layer. Improvements are however larger, when additionally updating the root-zone directly. For these two experiments, soil texture perturbations were reduced with increasing layer depth. CLM layer thickness ~~greatly~~ vastly increases with depth ~~which translates into identical increments in relative soil moisture being very different in~~ and homogeneous soil perturbations across all layers therefore result in large deep layer updates in terms of absolute soil moisture. ~~To thus restrict too large updates within the root zone we scaled the soil texture perturbations. The scaling factor for each layer was based on the ratio between layer thickness and the layer thickness of the top two layers, corresponding to the assumed depth were SMOS is sensitive, namely ca. 5 centimetres.~~ This was demonstrated in a third experiment, where correlation with in-situ measurements was highest compared to the first two experiments, namely 11 % and 8 % for top and root-zone soil moisture respectively.

Mean increments showed distinctive patterns with slight positive biases up to 1 % soil moisture in areas covered by ~~vegetation~~denser vegetation and neutral to slightly negative impact for areas mostly covered by sparse vegetation. Although this might be due to a problem with the uncalibrated forward operator~~or with the fixed high vegetation class LAI values within~~

the forward operator, a further possible cause is ~~,~~ it could also be caused by the use of climatological LAI data~~for CLM. The use of such data is common practise~~, which is common practice within current land data assimilation systems. ~~However, due~~ Due to the abundance of operational available vegetation data we would like to encourage future studies to look into possible improvements by using non-climatological LAI ~~. In fact, the climatological LAI data is often aggregated from monthly or sub-monthly LAI values and could be easily replaced with the non-aggregated data~~, where cloud cover permits. Climatological LAI might especially pose a problem for the monitoring of extreme events, such as droughts, since these ~~would likely~~ tend to result in lower LAI values again influencing the forward simulations. ~~With climatological LAI data these feedback processes will not be modelled. The potential of the direct brightness temperature assimilation, allowing for the consistent use of data across model and forward simulations, should be made full use of. Within assimilation systems where the forward operator is calibrated, it is also likely that the problem of equifinality for the forward operator parameters could thereby be reduced.~~

Long term assimilation effects were analysed by estimating the cumulative distribution functions for each grid cell prior to and after assimilation. On average, lower quantiles are shifted towards wetter conditions and higher quantiles are slightly shifted to drier conditions, ~~although the very high quantiles remained unchanged, overall resulting in a~~ resulting in reduced analysis variability. ~~This highlights the fact, that although in principle the assimilation experiment was set up using unbiased observations, analysis biases with non-linear behaviour may still be introduced. An additional experiment using~~ Spatial patterns in the quantiles change significantly at different quantile levels. We have shown these exemplary for the 10 % quantile. Here, for the experiment using homogeneous soil texture perturbations across all layers~~was carried out showing much larger analysis biases~~, the root-zone soil moisture showed a strong reduction compared to the ~~open loop~~ open-loop run at the 10 % quantile level. Patterns visible in the increment bias were strongly exaggerated, highlighting the problem of too large updates within the root zone and the general sensitivity towards model perturbations within the verticality.

~~The reduction of analysis variability by the assimilation might be partly attributed to the anomaly rescaling and to disentangle wished for systematic enhancements from erroneously introduced analysis is a challenge. There is still not the one perfect approach to rescale the observations to mach the model or to calibrate the forward observation and looking at long term CDF changes induced by the assimilation should be part of evaluating these different approaches, also keeping the final application of the analysis data in mind.~~

~~Longer~~ Operational land surface data assimilation systems are highly optimised, yet with longer L-band time series ~~are becoming available~~~~through the continuation of existing missions as well as new ones~~becoming available, such as ~~the Soil Moisture Active Passive Mission (SMAP, Entekhabi et al. (2010a)). This in the long run opens up more application possibilities, since soil moisture datasets enhanced by assimilating will allow for the computation of more stable grid cell CDFs. This will be especially useful within the context of hydrological~~ through the continuation of SMOS and SMAP, understanding the long-term effect of assimilating such data will gain importance. Draper and Reichle (2015) have for instance shown that data assimilation is able to correct modelled soil moisture also at longer time intervals from sub-seasonal to seasonal scale and seasonal differences in the assimilation effect are reported across many studies.

Hydrological monitoring systems, where it is important to identify the relative occurrence of certain soil moisture levels and to monitor patterns both over space and time. Within this study we have shown that the impact on CDFs will have an effect on the quantile based classification of a drought event and change the spatial extent of the affected area.

Especially for droughts these datasets are beneficial, since the assimilation can update the root zone soil moisture, which is of vital importance for plant growth and thus for monitoring agricultural drought. Many drought monitoring systemscurrently rely on precipitation based indices, e. g. the Palmer Drought Severity Index (PSDI, Palmer (1965)), due to the high correlation of soil moisture with precipitation over longer time spans. However,this correlation can be much smaller on small time spans and soil moisture is influenced by a multitude of factors within the land surface complex, which a land surface model is better able to represent across user-required time steps and spatial resolutions (Sheffield et al., 2004). But also in the case of floods an improved root-zone representation will be beneficial for predicting how much precipitation is required to saturate top layer soil.

On overall, within, are more and more likely to incorporate the assimilation of brightness temperatures with sufficiently long data records becoming available. Existing hydrological monitoring systems, such as the US. drought monitor (Svoboda et al., 2002), the African Flood and Drought Monitor (Sheffield et al., 2014), the German Drought Monitor (Samaniego et al., 2013) or the Australian Water Resource Assessment (Van Dijk et al., 2011; Vaze et al., 2013) all use soil moisture quantiles at grid cell level to characterise different levels of severity. Within this study we have attempted to show the possible complex behaviours induced by long-term assimilation and implications for the use of assimilation improved soil moisture simulations. Yet, it is clear that more studies should be carried out, bridging the gap between technical assimilation studies and the application. For instance, we have shown assimilation shown that the assimilation induced quantile changes for one specific data assimilation setup. The systematic analysis of various observation rescaling techniques and the impact on recorded drought and flood events should be part of future studieswill have an effect on the spatio-temporal classification of areas above or below a certain quantile level.

21

References

Balsamo, G., Beljaars, A., Scipal, K., Viterbo, P., van den Hurk, B., Hirschi, M., and Betts, A. K.: A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System, Journal of Hydrometeorology, 10, 623–643, 2009.

Barella-Ortiz, A., Polcher, J., de Rosnay, P., Piles, M., and Gelati, E.: Comparison of measured brightness temperatures from SMOS with modelled ones from ORCHIDEE and H-TESSEL over the Iberian Peninsula, Hydrology and Earth System Sciences Discussions, 12, 13 019–13 067, 2015.

Bonan, G. B., Levis, S., Kergoat, L., and Oleson, K. W.: Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models, Global Biogeochemical Cycles, 16, 5–1–5–23, 2002.

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

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

Champeaux, J. L., Masson, V., and Chauvin, F.: ECOCLIMAP: a global database of land surface parameters at 1 km resolution, Meteorological Applications, 12, 29–32, 2005.

Chen, F., Crow, W. T., and Ryu, D.: Dual forcing and state correction via soil moisture assimilation for improved rainfall–runoff modeling, Journal of Hydrometeorology, 15, 1832–1848, 2014.

Choudhury, B., Schmugge, T. J., Chang, A., and Newton, R.: Effect of surface roughness on the microwave emission from soils, Journal of Geophysical Research: Oceans, 84, 5699–5706, 1979.

De Lannoy, G. J., Reichle, R. H., Houser, P. R., Pauwels, V., and Verhoest, N. E.: Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter, Water Resources Research, 43, 2007.

De Lannoy, G. J. M. and Reichle, R. H.: Global assimilation of multiangle and multipolarization SMOS brightness temperature observations into the GEOS-5 Catchment Land Surface Model for soil moisture estimation, Journal of Hydrometeorology, 17, 669–691, 2016a.

De Lannoy, G. J. M. and Reichle, R. H.: Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model, Hydrology and Earth System Sciences, 20, 4895–4911, 2016b.

De Lannoy, G. J. M., Reichle, R. H., and Pauwels, V. R. N.: Global calibration of the GEOS-5 L-band microwave radiative transfer model over non frozen land using SMOS observations, Journal of Hydrometeorology, 14, 765–785, 2013.

de Rosnay, P., Drusch, M., Boone, A., Balsamo, G., Decharme, B., Harris, P., Kerr, Y., Pellarin, T., Polcher, J., and Wigneron, J.-P.: AMMA Land Surface Model Intercomparison Experiment coupled to the Community Microwave Emission Model: ALMIP-MEM, Journal of Geophysical Research: Atmospheres, 114, 2009.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the Royal Meteorological Society, 137, 553–597, 2011.

Dharssi, I., Bovis, K., Macpherson, B., and Jones, C.: Operational assimilation of ASCAT surface soil wetness at the Met Office, Hydrology and Earth System Sciences, 15, 2729–2746, 2011.

Dorigo, W., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., Oevelen, P. v., et al.: The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements, Hydrology and Earth System Sciences, 15, 1675–1698, 2011.

5   Draper, C. and Reichle, R.: The impact of near-surface soil moisture assimilation at subseasonal, seasonal, and inter-annual timescales, Hydrology and Earth System Sciences, 19, 4831–4844, 2015.

Draper, C., Mahfouf, J.-F., and Walker, J.: An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme, Journal of Geophysical Research: Atmospheres, 114, 2009.

Draper, C., Mahfouf, J.-F., Calvet, J.-C., Martin, E., and Wagner, W.: Assimilation of ASCAT near-surface soil moisture into the SIM

10   hydrological model over France, Hydrology and Earth System Sciences, 15, 3829, 2011.

Drusch, M., Holmes, T., de Rosnay, P., and Balsamo, G.: Comparing ERA-40-based L-band brightness temperatures with Skylab observations: A calibration/validation study using the Community Microwave Emission Model, Journal of Hydrometeorology, 10, 213, 2009.

Dumedah, G., Walker, J. P., and Merlin, O.: Root-zone soil moisture estimation from assimilation of downscaled Soil Moisture and Ocean Salinity data, Advances in Water Resources, 84, 14 – 22, 2015.

15   Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., Kimball, J., Piepmeier, J. R., Koster, R. D., Martin, N., McDonald, K. C., Moghaddam, M., Moran, S., Reichle, R., Shi, J. C., Spencer, M. W., Thurman, S. W., Tsang, L., and Zyl, J. V.: The Soil Moisture Active Passive (SMAP) Mission, Proceedings of the IEEE, 98, 704–716, 2010a.

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J.,

20   Johnson, J., et al.: The soil moisture active passive (SMAP) mission, Proceedings of the IEEE, 98, 704–716, 2010b.

Han, X., Li, X., Hendricks Franssen, H., Vereecken, H., and Montzka, C.: Spatial horizontal correlation characteristics in the land data assimilation of soil moisture, Hydrology and Earth System Sciences, 16, 1349–1363, 2012.

Han, X., Hendricks Franssen, H.-J., Li, X., Zhang, Y., Montzka, C., and Vereecken, H.: Joint assimilation of surface temperature and L-band microwave brightness temperature in land data assimilation, Vadose Zone Journal, 12, 2013.

25   Han, X., Franssen, H.-J. H., Montzka, C., and Vereecken, H.: Soil moisture and soil properties estimation in the Community Land Model with synthetic brightness temperature observations, Water Resources Research, 50, 6081–6105, 2014a.

Han, X., Franssen, H.-J. H., Montzka, C., and Vereecken, H.: Soil moisture and soil properties estimation in the Community Land Model with synthetic brightness temperature observations, Water resources research, 50, 6081–6105, 2014b.

Han, X., Li, X., He, G., Kumbhar, P., Montzka, C., Kollet, S., Miyoshi, T., Rosolem, R., Zhang, Y., Vereecken, H., and Franssen, H.-J. H.:

30   DasPy 1.0 - the Open Source Multivariate Land Data Assimilation Framework in combination with the Community Land Model 4.5, Geoscientific Model Development Discussions, 8, 7395–7444, 2015a.

Han, X., Li, X., He, G., Kumbhar, P., Montzka, C., Kollet, S., Miyoshi, T., Rosolem, R., Zhang, Y., Vereecken, H., et al.: DasPy 1.0–the Open Source Multivariate Land Data Assimilation Framework in combination with the Community Land Model 4.5, Geoscientific Model Development Discussions, 8, 2015b.

35   Han, X., Li, X., Rigon, R., Jin, R., and Endrizzi, S.: Soil moisture estimation by assimilating L-band Microwave brightness temperature with geostatistics and observation localization, PloS one, 10, e0116 435, 2015c.

Hawdon, A., McJannet, D., and Wallace, J.: Calibration and correction procedures for cosmic-ray neutron soil moisture probes located across Australia, Water Resources Research, 50, 5029–5043, 2014.

Hengl, T., de Jesus, J. M., MacMillan, R., Batjes, N. H., Heuvelink, G. B., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J. G., Walsh, M. G., and Gonzalez, M. R.: SoilGrids1km - Global Soil Information Based on Automated Mapping, PLoS ONE, 9, 2014.

Herrera-Estrada, J. E., Satoh, Y., and Sheffield, J.: Spatiotemporal dynamics of global drought, Geophysical Research Letters, pp. n/a–n/a, 2016GL071768, 2017.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas, International Journal of Climatology, 25, 1965–1978, 2005.

Hornbuckle, B. K. and England, A. W.: Diurnal variation of vertical temperature gradients within a field of maize: Implications for satellite microwave radiometry, IEEE Geoscience and Remote Sensing Letters, 2, 74–77, 2005.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D: Nonlinear Phenomena, 230, 112 – 126, data Assimilation, 2007.

Jia, B., Xie, Z., Tian, X., and Shi, C.: A soil moisture assimilation scheme based on the ensemble Kalman filter using microwave brightness temperature, Science in China Series D: Earth Sciences, 52, 1835–1848, 2009.

Johnson, F., White, C. J., van Dijk, A., Ekstrom, M., Evans, J. P., Jakob, D., Kiem, A. S., Leonard, M., Rouillard, A., and Westra, S.: Natural hazards in Australia: floods, Climatic Change, 139, 21–35, 2016.

Kalman, R. E. et al.: A new approach to linear filtering and prediction problems, Journal of basic Engineering, 82, 35–45, 1960.

Ke, Y., Leung, L. R., Huang, M., Coleman, A. M., Li, H., and Wigmosta, M. S.: Development of high resolution land surface parameters for the Community Land Model, Geoscientific Model Development, 5, 1341–1362, 2012.

Kerr, Y. H., Waldteufel, P., Wigneron, J. P., Martinuzzi, J., Font, J., and Berger, M.: Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission, IEEE Transactions on Geoscience and Remote Sensing, 39, 1729–1735, 2001.

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N., Gruhier, C., et al.: The SMOS mission: New tool for monitoring key elements ofthe global water cycle, Proceedings of the IEEE, 98, 666–687, 2010.

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Bitar, A. A., Cabot, F., Gruhier, C., Juglea, S. E., Leroux, D., Mialon, A., and Delwart, S.: The SMOS Soil Moisture Retrieval Algorithm, IEEE Transactions on Geoscience and Remote Sensing, 50, 1384–1403, 2012.

Kiem, A. S., Johnson, F., Westra, S., van Dijk, A., Evans, J. P., O'Donnell, A., Rouillard, A., Barr, C., Tyler, J., Thyer, M., Jakob, D., Woldemeskel, F., Sivakumar, B., and Mehrotra, R.: Natural hazards in Australia: droughts, Climatic Change, 139, 37–54, 2016.

Kumar, S. V., Reichle, R. H., Koster, R. D., Crow, W. T., and Peters-Lidard, C. D.: Role of Subsurface Physics in the Assimilation of Surface Soil Moisture Observations, Journal of Hydrometeorology, 10, 1534–1547, 2009.

Lehner, B., Verdin, K., and Jarvis, A.: New global hydrography derived from spaceborne elevation data, Eos, 89, 93–94, 2008.

Leroux, D. J., Kerr, Y. H., Richaume, P., and Fieuzal, R.: Spatial distribution and possible sources of SMOS errors at the global scale, Remote Sensing of Environment, 133, 240 – 250, 2013.

Leroux, D. J., Pellarin, T., Vischel, T., Cohard, J.-M., Gascon, T., Gibon, F., Mialon, A., Galle, S., Peugeot, C., and Seguis, L.: Assimilation of SMOS soil moisture into a distributed hydrological model and impacts on the water cycle variables over the Ouémé catchment in Benin, Hydrology and Earth System Sciences, 20, 2827–2840, 2016.

Lievens, H., Al Bitar, A., Verhoest, N., Cabot, F., De Lannoy, G., Drusch, M., Dumedah, G., Franssen, H., Kerr, Y., Tomer, S., Martens, B., Merlin, O., Pan, M., van den Berg, M., Vereecken, H., Walker, J., Wood, E., and Pauwels, V.: Optimization of a radiative transfer forward operator for simulating SMOS brightness temperatures over the Upper Mississippi Basin, Journal of Hydrometeorology, 16, 1109–1134, 2015a.

Lievens, H., Tomer, S., Al Bitar, A., De Lannoy, G., Drusch, M., Dumedah, G., Hendricks Franssen, H.-J., Kerr, Y., Martens, B., Pan, M., Roundy, J., Vereecken, H., Walker, J., Wood, E., Verhoest, N., and Pauwels, V.: SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia, Remote Sensing of Environment, 168, 146–162, 2015b.

5 Lievens, H., Tomer, S. K., Al Bitar, A., De Lannoy, G., Drusch, M., Dumedah, G., Franssen, H.-J. H., Kerr, Y., Martens, B., Pan, M., et al.: SMOS soil moisture assimilation for improved hydrologic simulation in the Murray Darling Basin, Australia, Remote Sensing of Environment, 168, 146–162, 2015c.

Lievens, H., De Lannoy, G., Al Bitar, A., Drusch, M., Dumedah, G., Hendricks Franssen, H.-J., Kerr, Y., Tomer, S., Martens, B., Merlin, O., Pan, M., Roundy, J., Vereecken, H., Walker, J., Wood, E., Verhoest, N., and Pauwels, V.: Assimilation of SMOS soil moisture and

10 brightness temperature products into a land surface model, Remote Sensing of Environment, 180, 292–304, 2016.

Martens, B., Miralles, D., Lievens, H., Fernández-Prieto, D., and Verhoest, N.: Improving terrestrial evaporation estimates over continental Australia through assimilation of SMOS soil moisture, International Journal of Applied Earth Observation and Geoinformation, 48, 146–162, 2016a.

Martens, B., Miralles, D., Lievens, H., Fernández-Prieto, D., and Verhoest, N. E.: Improving terrestrial evaporation estimates over continental

15 Australia through assimilation of SMOS soil moisture, International Journal of Applied Earth Observation and Geoinformation, 48, 146–162, 2016b.

McColl, K., Pipunic, R., Ryu, D., and Walker, J.: Validation of the MODIS LAI product in the Murrumbidgee catchment, Australia, in: Proc. 19th Int. Congress on Modelling and Simulation, pp. 1973–1979, 2011.

Mecklenburg, S., Drusch, M., Kaleschke, L., Rodriguez-Fernandez, N., Reul, N., Kerr, Y., Font, J., Martin-Neira, M., Oliva, R., Daganzo-

20 Eusebio, E., Grant, J., Sabia, R., Macelloni, G., Rautiainen, K., Fauste, J., de Rosnay, P., Munoz-Sabater, J., Verhoest, N., Lievens, H., Delwart, S., Crapolicchio, R., de la Fuente, A., and Kornberg, M.: ESA's Soil Moisture and Ocean Salinity mission: from science to operational applications, Remote Sensing of Environment, 180, 3–18, 2016.

Merlin, O., Rudiger, C., Al Bitar, A., Richaume, P., Walker, J. P., and Kerr, Y. H.: Disaggregation of SMOS soil moisture in Southeastern Australia, IEEE Transactions on Geoscience and Remote Sensing, 50, 1556–1571, 2012.

25 Mironov, V. L., Dobson, M. C., Kaupp, V. H., Komarov, S. A., and Kleshchenko, V. N.: Generalized refractive mixing dielectric model for moist soils, IEEE Transactions on Geoscience and Remote Sensing, 42, 773–785, 2004.

Miyoshi, T. and Yamane, S.: Local Ensemble Transform Kalman Filtering with an AGCM at a T159/L48 Resolution, Monthly Weather Review, 135, 3841–3861, 2007.

Mohanty, B. P., Cosh, M. H., Lakshmi, V., and Montzka, C.: Soil Moisture Remote Sensing: State-of-the-Science, Vadose Zone Journal, 16,

30 2017.

Montaldo, N., Albertson, J. D., Mancini, M., and Kiely, G.: Robust simulation of root zone soil moisture with assimilation of surface soil moisture data, Water Resources Research, 37, 2889–2900, 2001.

Montzka, C., Moradkhani, H., Weihermüller, L., Franssen, H.-J. H., Canty, M., and Vereecken, H.: Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter, Journal of Hydrology, 399, 410–421, 2011.

35 Montzka, C., Pauwels, V., Franssen, H.-J. H., Han, X., and Vereecken, H.: Multivariate and multiscale data assimilation in terrestrial systems: A review, Sensors, 12, 16 291–16 333, 2012.

Muñoz-Sabater, J.: Incorporation of passive microwave Brightness Temperatures in the ECMWF soil moisture analysis, Remote Sensing, 7, 5758–5784, 2015.

Muñoz-Sabater, J., Fouilloux, A., and de Rosnay, P.: Technical implementation of SMOS data in the ECMWF Integrated Forecasting System, IEEE Geoscience and Remote Sensing Letters, 9, 252–256, 2012.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, W. J., Subin, Z. M., Swenson,
5   S. C., and Thornton, P. E.: Technical Description of version 4.5 of the Community Land Model (CLM), Tech. rep., 2013.

Palmer, W. C.: Meteorological drought, vol. 30, US Department of Commerce, Weather Bureau Washington, DC, 1965.

Panciera, R., Walker, J. P., Kalma, J. D., Kim, E. J., Hacker, J. M., Merlin, O., Berger, M., and Skou, N.: The NAFE'05/CoSMOS Data Set: Toward SMOS Soil Moisture Retrieval, Downscaling, and Assimilation, IEEE Transactions on Geoscience and Remote Sensing, 46, 736–745, 2008.

10 Panciera, R., Walker, J. P., Jackson, T. J., Gray, D. A., Tanase, M. A., Ryu, D., Monerris, A., Yardley, H., Rüdiger, C., Wu, X., Gao, Y., and Hacker, J. M.: The Soil Moisture Active Passive Experiments (SMAPEx): Toward Soil Moisture Retrieval From the SMAP Mission, IEEE Transactions on Geoscience and Remote Sensing, 52, 490–507, 2014.

Parada, L. M. and Liang, X.: Optimal multiscale Kalman filter for assimilation of near-surface soil moisture into land surface models, Journal of Geophysical Research: Atmospheres, 109, 2004.

15 Peischl, S., Walker, J., Allahmoradi, M., Barrett, D., Gurney, R., Kerr, Y., Kim, E., Le Marshall, J., Rüdiger, C., Ryu, D., et al.: Towards validation of SMOS using airborne and ground data over the Murrumbidgee catchment, in: Proc. MODSIM, pp. 3733–3739, 2009.

Peischl, S., Walker, J. P., Rüdiger, C., Ye, N., Kerr, Y. H., Kim, E., Bandara, R., and Allahmoradi, M.: The AACES field experiments: SMOS calibration and validation across the Murrumbidgee River catchment, Hydrology and Earth System Sciences, 16, 1697–1708, 2012.

Pellarin, T., Wigneron, J. P., Calvet, J. C., Berger, M., Douville, H., Ferrazzoli, P., Kerr, Y. H., Lopez-Baeza, E., Pulliainen, J., Simmonds,
20   L. P., and Waldteufel, P.: Two-year global simulation of L-band brightness temperatures over land, IEEE Transactions on Geoscience and Remote Sensing, 41, 2135–2139, 2003.

Piles, M., Camps, A., Vall-Llossera, M., Corbella, I., Panciera, R., Rudiger, C., Kerr, Y. H., and Walker, J.: Downscaling SMOS-derived soil moisture using MODIS visible/infrared data, IEEE Transactions on Geoscience and Remote Sensing, 49, 3156–3166, 2011.

Reichle, R. H.: Data assimilation methods in the Earth sciences, Advances in Water Resources, 31, 1411–1418, 2008.

25 Reichle, R. H., Koster, R. D., Liu, P., Mahanama, S. P., Njoku, E. G., and Owe, M.: 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: Atmospheres, 112, 2007.

Renzullo, L. J., Van Dijk, A., Perraud, J.-M., Collins, D., Henderson, B., Jin, H., Smith, A., and McJannet, D.: Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment, Journal of hydrology, 519, 2747–2762,
30   2014.

Ridler, M.-E., Madsen, H., Stisen, S., Bircher, S., and Fensholt, R.: Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark, Water Resources Research, 50, 8962–8981, 2014.

Samaniego, L., Kumar, R., and Zink, M.: Implications of parameter uncertainty on soil moisture drought analysis in Germany, Journal of Hydrometeorology, 14, 47–68, 2013.

35 Scholze, M., Kaminski, T., Knorr, W., Blessing, S., Vossbeck, M., Grant, J., and Scipal, K.: Simultaneous assimilation of SMOS soil moisture and atmospheric CO2 in-situ observations to constrain the global terrestrial carbon cycle, Remote Sensing of Environment, 180, 334 – 345, special Issue: ESA's Soil Moisture and Ocean Salinity Mission - Achievements and Applications, 2016.

Sheffield, J., Goteti, G., Wen, F., and Wood, E. F.: A simulated soil moisture based drought analysis for the United States, Journal of Geophysical Research: Atmospheres, 109, n/a–n/a, d24108, 2004.

Sheffield, J., Wood, E. F., Chaney, N., Guan, K., Sadri, S., Yuan, X., Olang, L., Amani, A., Ali, A., Demuth, S., et al.: A drought monitoring and forecasting system for sub-Sahara African water resources and food security, Bulletin of the American Meteorological Society, 95, 861–882, 2014.

5 Smith, A. B., Walker, J. P., Western, A. W., Young, R. I., Ellett, K. M., Pipunic, R. C., Grayson, R. B., Siriwardena, L., Chiew, F. H. S., and Richter, H.: The Murrumbidgee soil moisture monitoring network data set, Water Resources Research, 48, 2012.

Su, C.-H., Ryu, D., Young, R. I., Western, A. W., and Wagner, W.: Inter-comparison of microwave satellite soil moisture retrievals over the Murrumbidgee Basin, southeast Australia, Remote Sensing of Environment, 134, 1 – 11, 2013.

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., et al.: The
10 drought monitor, Bulletin of the American Meteorological Society, 83, 1181–1190, 2002.

Tian, S., Tregoning, P., Renzullo, L. J., van Dijk, A. I. J. M., Walker, J. P., Pauwels, V. R. N., and Allgeyer, S.: Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals, Water Resources Research, 53, 1820–1840, 2017.

Van der Schalie, R., Parinussa, R., Renzullo, L. J., Van Dijk, A., Su, C.-H., and de Jeu, R. A.: SMOS soil moisture retrievals using the land
15 parameter retrieval model: Evaluation over the Murrumbidgee Catchment, southeast Australia, Remote Sensing of Environment, 163, 70–79, 2015.

Van Dijk, A. and Renzullo, L. J.: Water resource monitoring systems and the role of satellite observations, Hydrology and Earth System Sciences, 15, 39–55, 2011.

Van Dijk, A., Bacon, D., Barratt, D., Crosbie, R., Daamen, C., Fitch, P., Frost, A., Guerschman, J., Henderson, B., King, E., et al.: Design
20 and development of the Australian Water Resources Assessment system, in: Proceedings, Water Information Research and Development Alliance Science Symposium, 2011.

Vaze, J., Viney, N., Stenson, M., Renzullo, L., Van Dijk, A., Dutta, D., Crosbie, R., Lerat, J., Penton, D., Vleeshouwer, J., et al.: The Australian Water Resource Assessment Modelling System (AWRA), in: 20th International Congress on Modelling and Simulation, Adelaide, Australia, vol. 16, 2013.

25 Vereecken, H., Schnepf, A., Hopmans, J., Javaux, M., Or, D., Roose, T., Vanderborght, J., Young, M., Amelung, W., Aitkenhead, M., et al.: Modeling soil processes: Review, key challenges, and new perspectives, Vadose Zone Journal, 15, 2016.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resources Research, 50, 7505–7514, 2014.

Wigneron, J.-P., Laguerre, L., and Kerr, Y. H.: A simple parameterization of the L-band microwave emission from rough agricultural soils,
30 IEEE Transactions on Geoscience and Remote Sensing, 39, 1697–1707, 2001.

Wigneron, J.-P., Kerr, Y., Waldteufel, P., Saleh, K., Escorihuela, M.-J., Richaume, P., Ferrazzoli, P., De Rosnay, P., Gurney, R., Calvet, J.-C., et al.: L-band microwave emission of the biosphere (L-MEB) model: Description and calibration against experimental data sets over crop fields, Remote Sensing of Environment, 107, 639–655, 2007.

Yang, K., Watanabe, T., Koike, T., Li, X., Fujii, H., Tamagawa, K., Ishikawa, H., et al.: Auto-calibration system developed to assimilate
35 AMSR-E data into a land surface model for estimating soil moisture and the surface energy budget, Journal of the Meteorological Society of Japan. Ser. II, 85, 229–242, 2007.

Yilmaz, M. T. and Crow, W. T.: The optimality of potential rescaling approaches in land data assimilation, Journal of Hydrometeorology, 14,
910 650–660, 2013.

Table 1. CLM soil layer depths and relative layer thickness in respect to sum of the two top layers. The relative thickness was used as a scaling factor for the soil perturbations, effectively decreasing ensemble spread and error covariance for lower levels.

Layer Depth [m]	0.018	0.045	0.09	0.17	0.290	0.493	0.829	1.383	2.296	3.802
Perturb. scaling	1	1	1	0.60	0.36	0.22	0.13	0.08	0.05	0.03
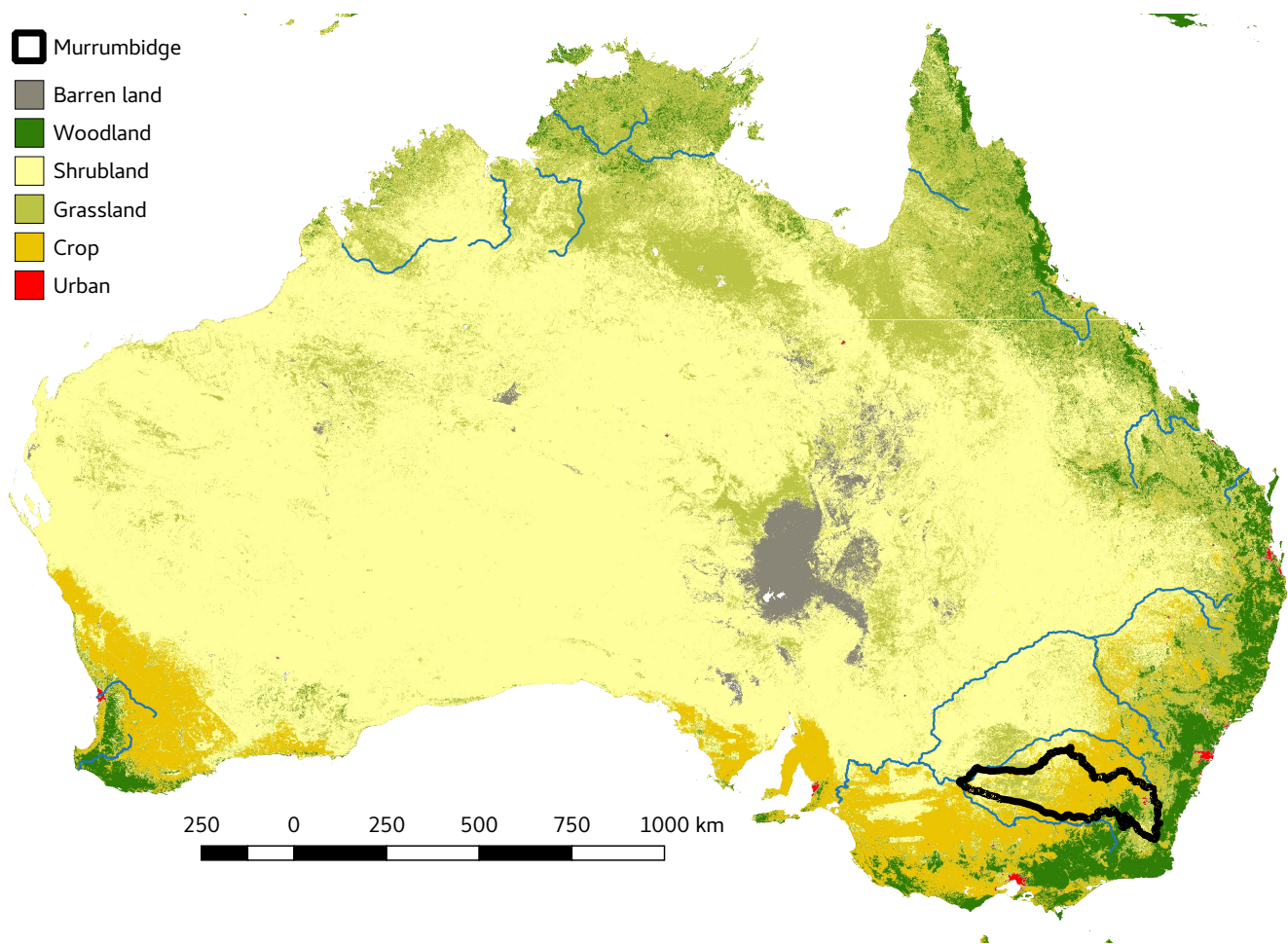
Figure 1. ~~Taylor diagram showing assimilation impact~~ CLM Plant Functional Types based on ~~top layer soil moisture, defined as 8 cm soil depth, (left)~~ MODIS MCDQ12 land cover classification and ~~lower level soil moisture (right) in terms of correlation coefficient R~~ECOCLIMAP climate zones at 500 m resolution prior to the aggregation to 0.25 degrees. Some classes are here aggregated for visualisation purpose, ~~standard deviation~~ e.g. evergreen temperate and ~~normalised RMSE~~evergreen tropical forests are both shown as Woodland. The ~~colours correspond to~~ boundary of the Murrumbidgee catchment, which is the site of the OzNet in-situ ~~measurement sites but are not comparable between both panels~~measurements, is shown.
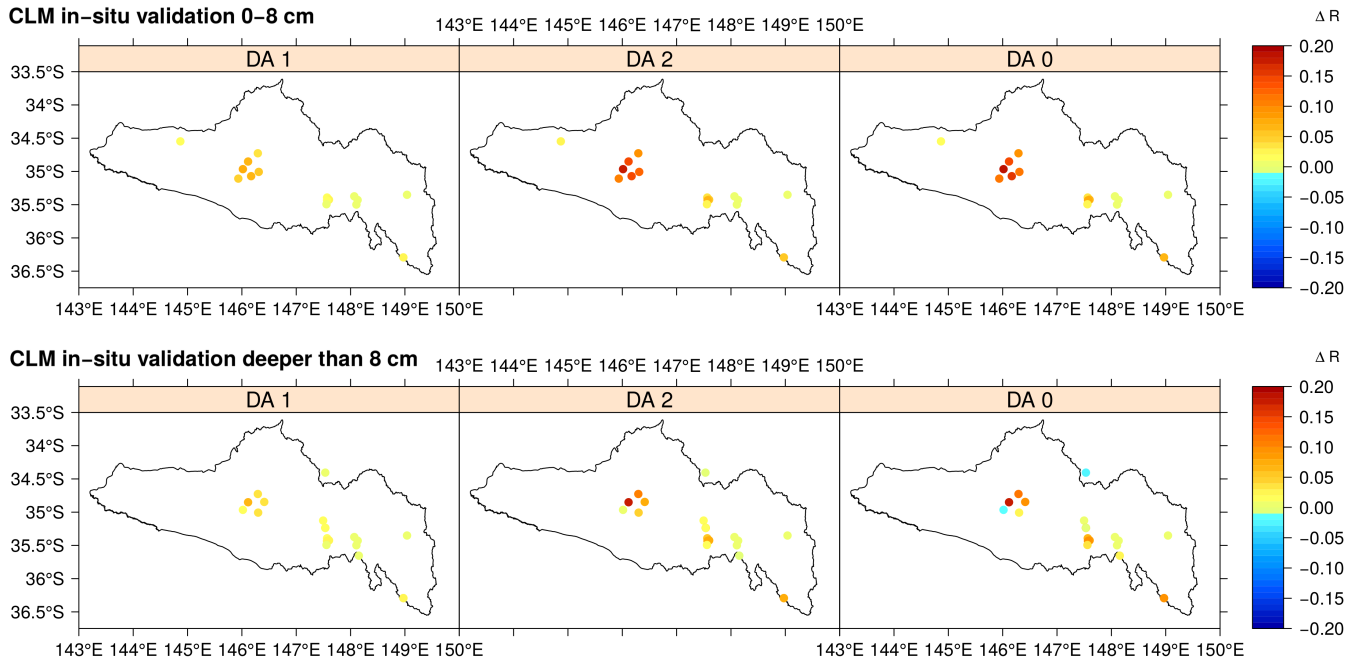
Figure 2. Change in correlation R for experiments DA 1, DA 2 and DA 0 both for top layer soil moisture (top panel) as well as the root-zone soil moisture (bottom panel) within the Murrumbidgee catchment. In the case of multiple measurements at the same location, the weighted average of the measured as well as modelled soil moisture was computed in accordance to the corresponding CLM layer thickness.

31
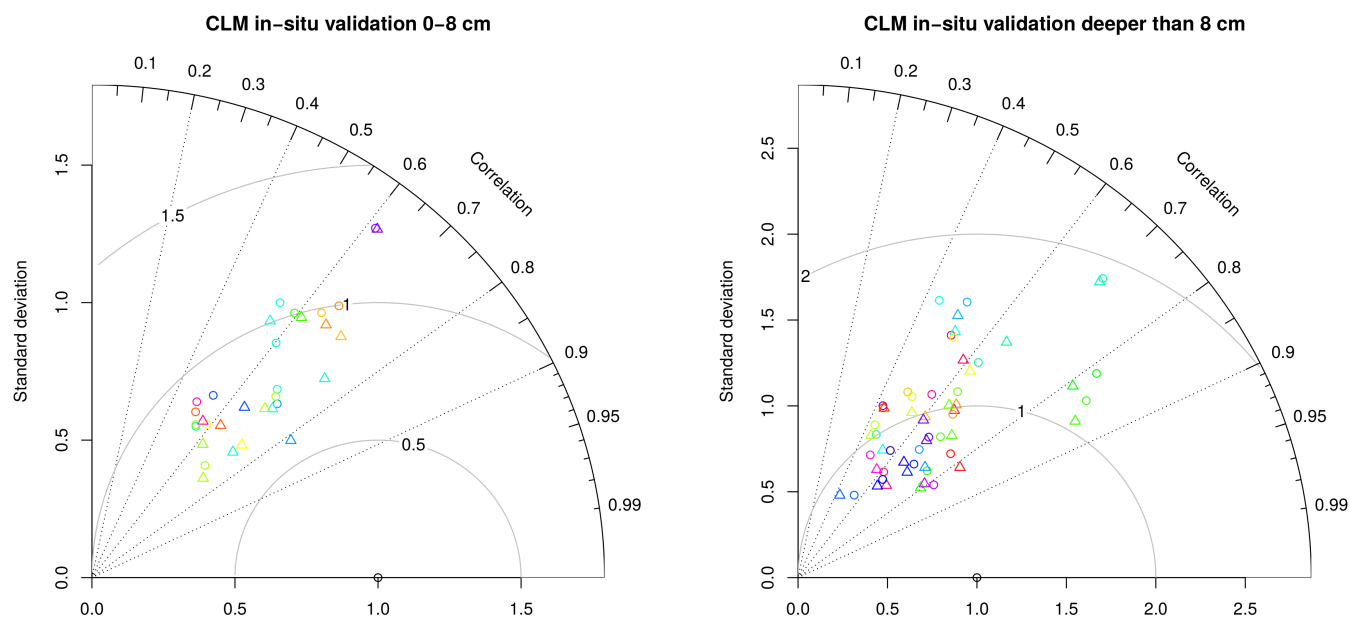
Figure 3. Taylor diagram showing assimilation impact on top layer soil moisture, defined as 8 cm soil depth, (left) and lower level soil moisture (right) in terms of correlation coefficient R, standard deviation and normalised RMSE.
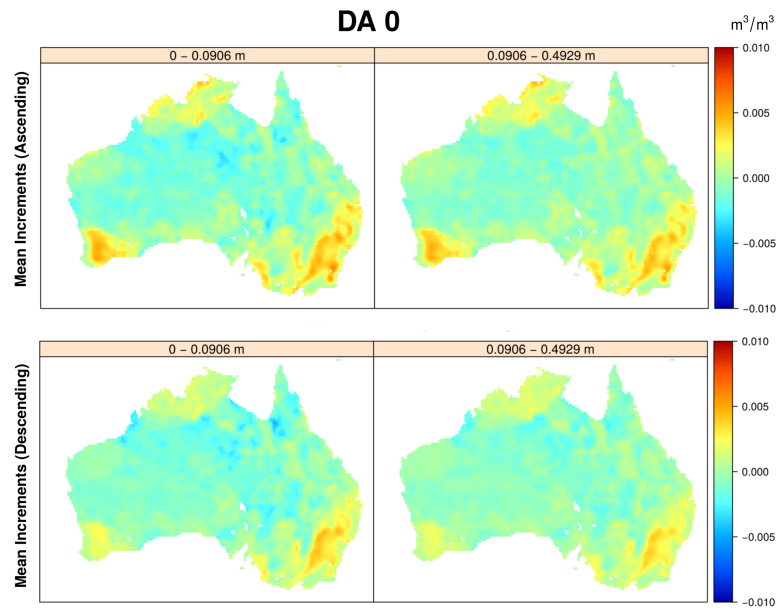
Figure 4. Mean of all increments for experiment DA 0 for top-layer soil moisture (left) and root zone soil moisture (right) for ascending (above) and descending orbits (below).
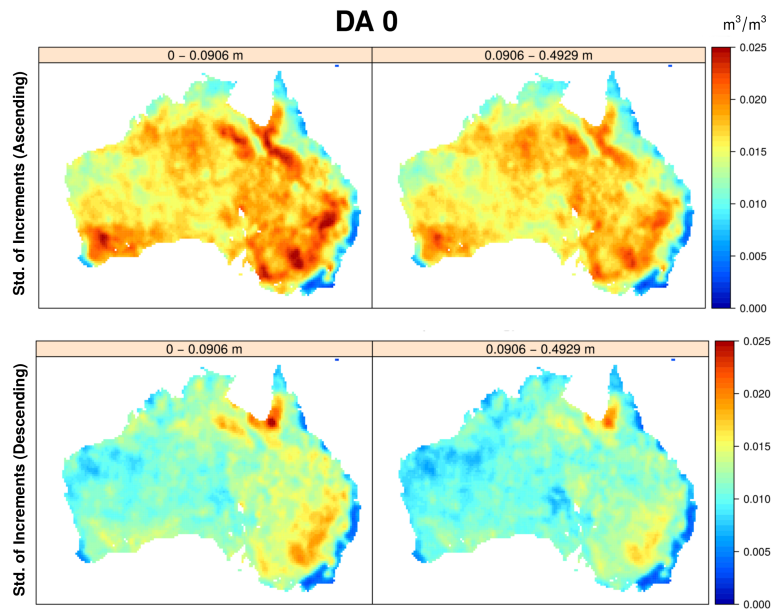
Figure 5. Standard deviation of all increments for experiment DA 0 for top-layer soil moisture (left) and root-zone soil moisture (right) for ascending (above) and descending orbits (below). Increments for the root-zone soil moisture are fairly similar to the top soil layers, due to the homogeneous texture perturbations applied across all layers.
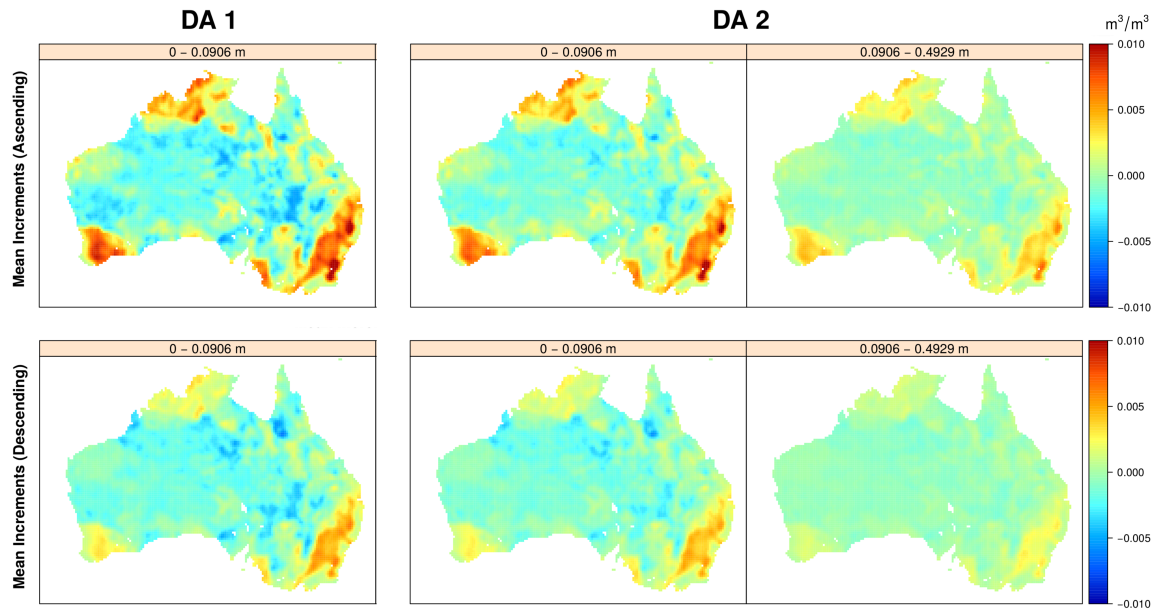
Figure 6. Mean of all increments for experiment DA 1 / DA 2 for top-layer soil moisture and root zone soil moisture for ascending (above) and descending orbits (below). Biases are strongest for the ascending orbit and distinctive spatial patterns are visible. Biases are strongly reduced both for deeper layers and descending orbits.
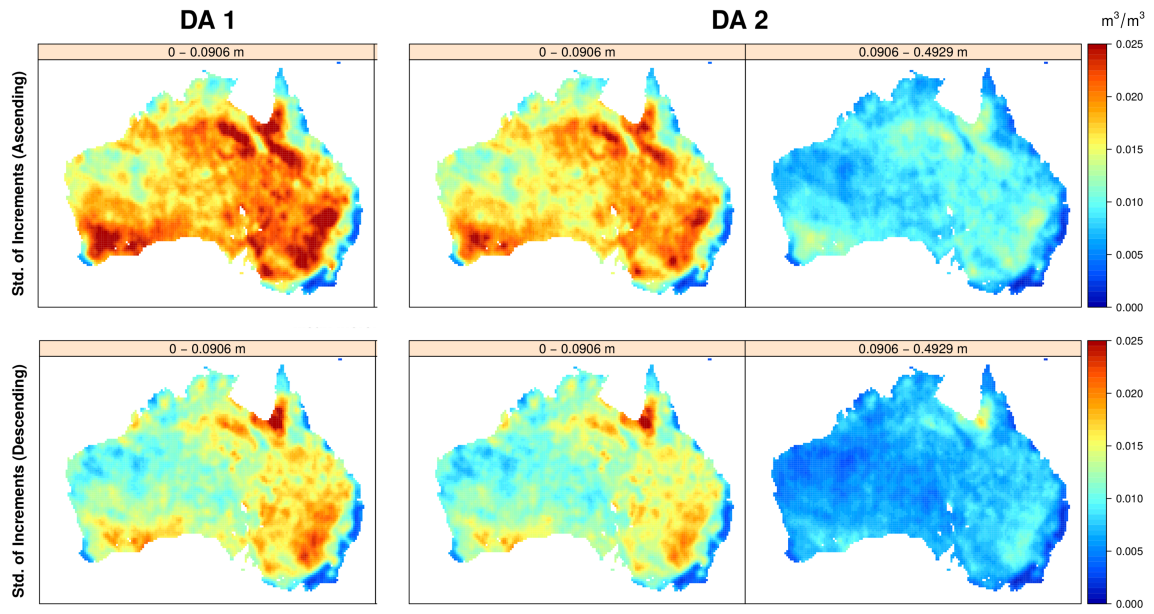
Figure 7. Standard deviation of all increments for experiment DA 1 / DA 2 for top-layer soil moisture and root zone soil moisture for ascending (above) and descending orbits (below). Increments are strongest for the ascending orbit and for top-layer soil moisture and even stronger when restricting assimilation to these layers, as in DA 1. Increments are very low or zero for the forested areas along the coastline, either due to the absence of observations or the high LAI values masking any soil moisture signal within the forward operator.
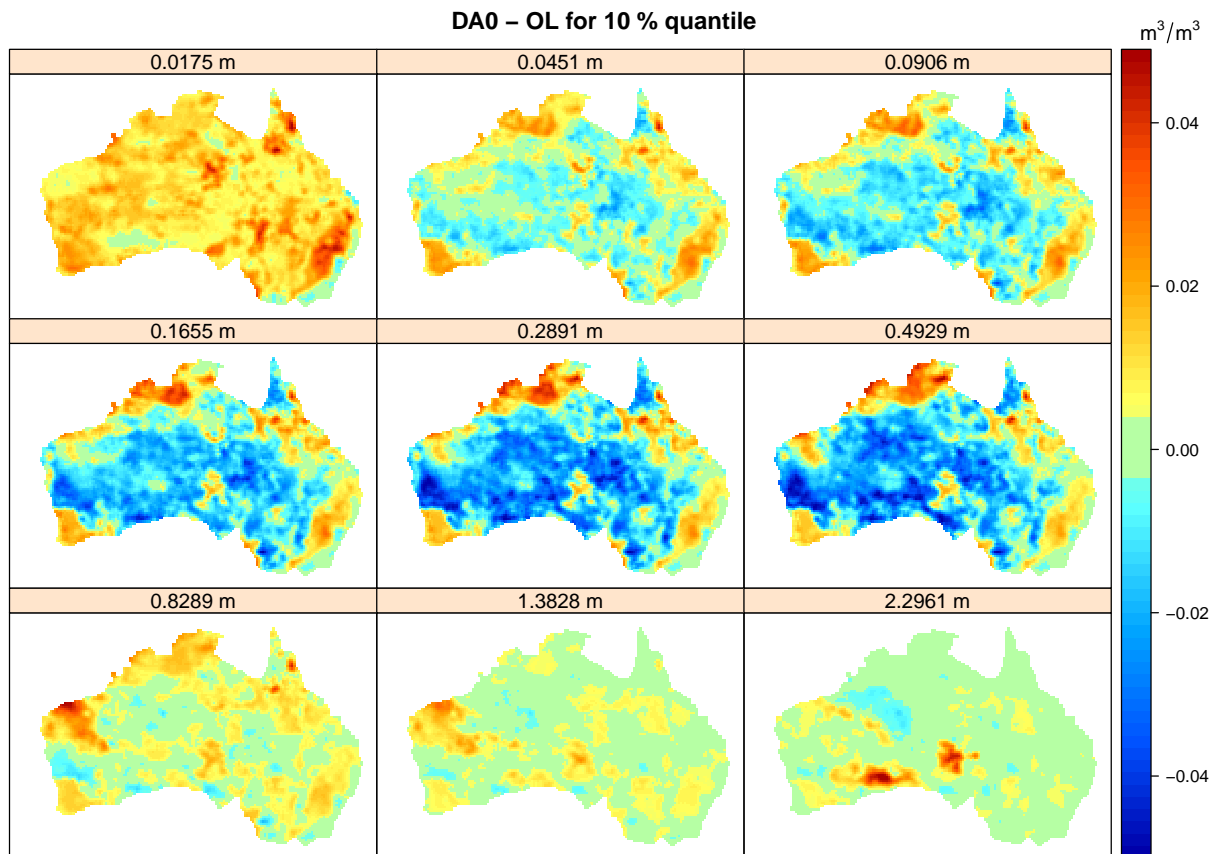
Figure 8. Differences in relative soil moisture ~~%/100~~between ~~open loop~~ open-loop and ~~DA0~~ DA 0 experiment (DA 0 - OL) for the 10 % quantile. The individual panels correspond to the top 9 CLM soil layers.
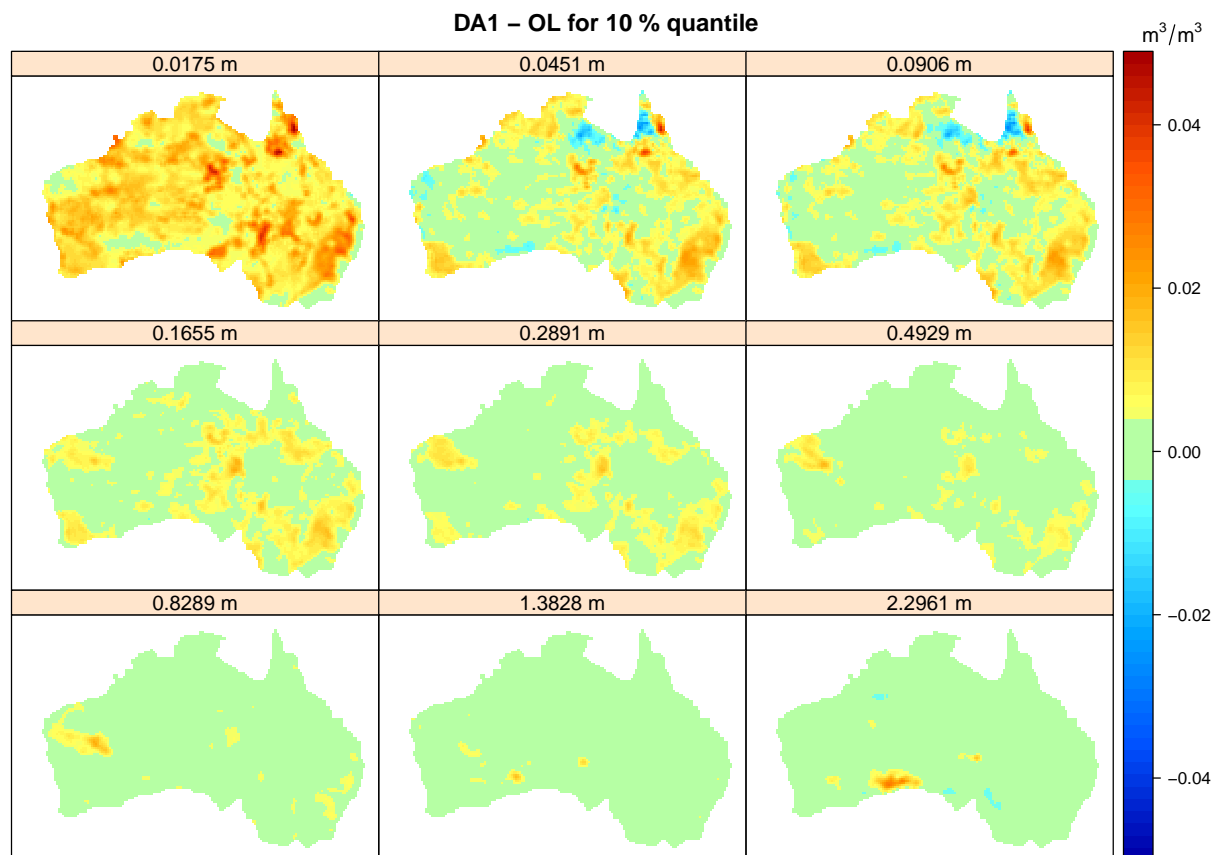
Figure 9. Differences in relative soil moisture %/100between open loop open-loop and DA1 DA 1 experiment (DA 1 - OL) for 10 % quantile. The individual panels correspond to the top 9 CLM soil layers titled with their depth.

**DA2 – OL for 10 % quantile**
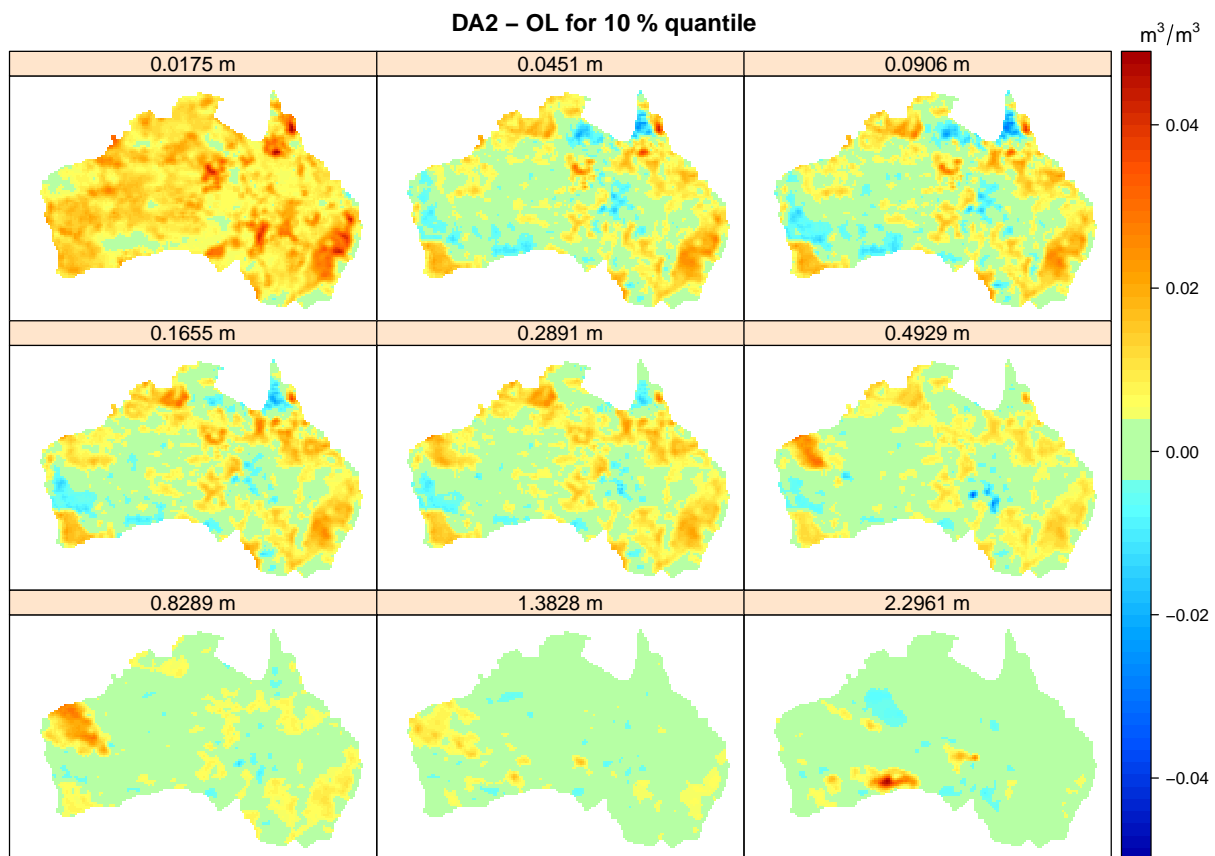
Figure 10. Differences in relative soil moisture ~~%/100~~between ~~open loop~~ open-loop and ~~DA2~~ DA 2 experiment (~~DA2~~ DA 2 - OL) for the 10 % quantile. The individual panels correspond to the top 9 CLM soil layers.
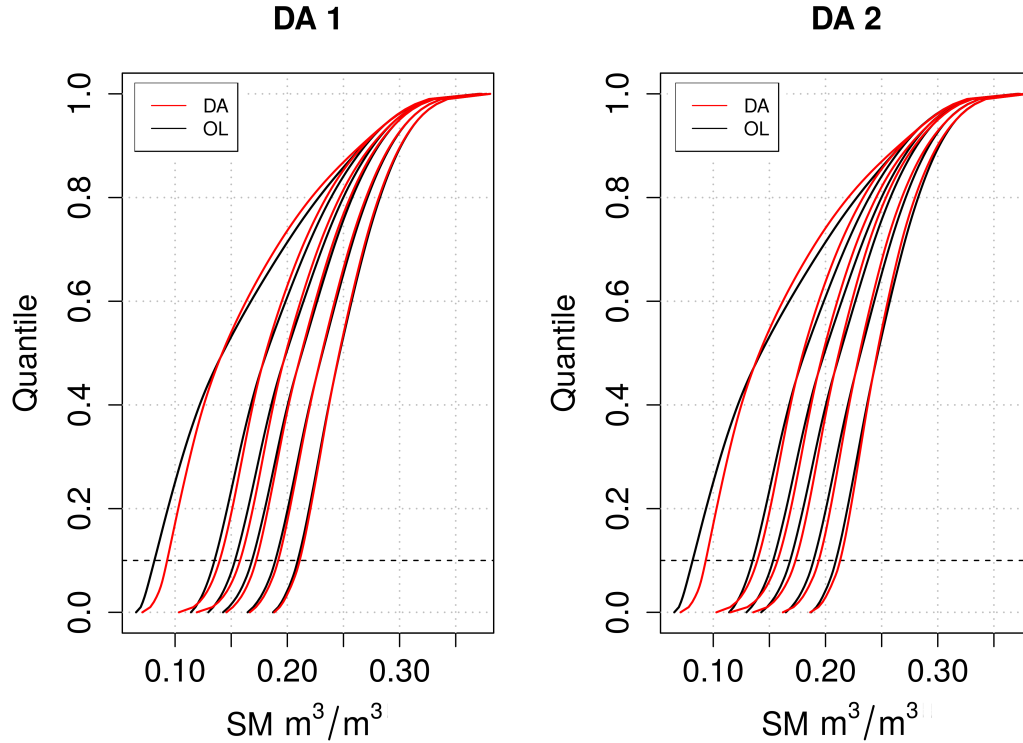
Figure 11. Cumulative distribution functions (CDFs) for the upper 6 CLM soil layers for experiments ~~DA1~~ DA 1 and ~~DA2~~ DA 2, based on quantiles computed for all data across the model domain. CDFs for ~~open loop~~ open-loop simulations are shown in black and assimilation results in red. Both panels show changes in CDF behaviour for the layers being updated in the respective experiments, i.e. layers 1-3 for ~~DA1~~ DA 1 and layers 1-6 for ~~DA2.~~ DA 2. Soil moisture increases systematically with soil depth allowing for the easy identification of the layers within the plot. The dashed vertical line marks the 10 % quantile, corresponding to figures 9 and 10.
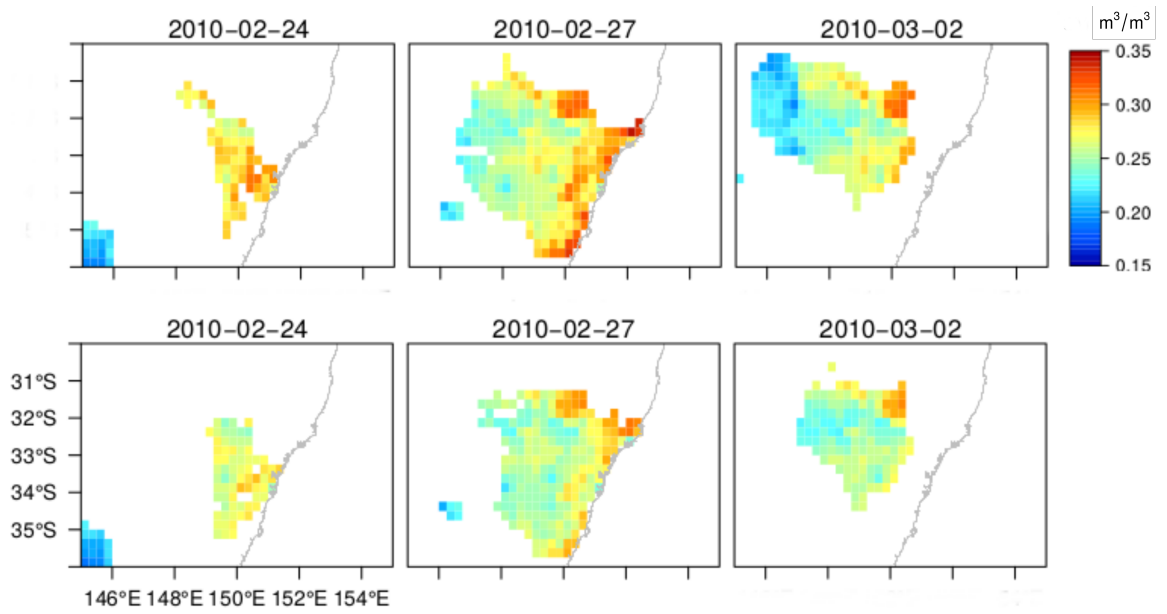
Figure 12. Sample drought event for February 2010, showing only the root zone soil moisture below the 10 % quantile level for the ~~open loop~~ open-loop (above) and experiment ~~DA2~~ DA 2 (below). The different spatial extent and differences in soil moisture itself, depending on the dataset used, at three different days are clearly visible. The figure is centred around the Central coast of New South Wales.