Uncertainty Quantification of GEOS-5 L-Band Radiative Transfer Model Parameters using Bayesian Inference and SMOS Observations

Gabriëlle J. M. De Lannoy[a,b], Rolf H. Reichle[a], Jasper A. Vrugt[c,d,e]

[a] *NASA Goddard Space Flight Center, Code 610.1, Greenbelt, MD 20771, USA*
[b] *GESTAR, Universities Space Research Association, Columbia, MD 21044, USA*
[c] *University of California, Irvine, Department of Civil and Environmental Engineering, Irvine, CA 92697-2175, USA*
[d] *University of California, Irvine, Department of Earth System Science, Irvine, CA 92697-2175, USA*
[e] *Institute for Biodiversity and Ecosystems Dynamics, University of Amsterdam, 1098 XH Amsterdam, The Netherlands*

Abstract

Uncertainties in L-band (1.4 GHz) radiative transfer modeling (RTM) affect the simulation of brightness temperatures (Tb) over land and the inversion of satellite-observed Tb into soil moisture retrievals. In particular, accurate estimates of the microwave soil roughness, vegetation opacity and scattering albedo for large-scale applications are difficult to obtain from field studies and often lack an uncertainty estimate. Here, a Markov Chain Monte Carlo (MCMC) simulation method is used to determine satellite-scale estimates of RTM parameters and their posterior uncertainty by minimizing the misfit between long-term averages and standard deviations of simulated and observed Tb at a range of incidence angles, at horizontal and vertical polarization, and for morning and evening overpasses. Tb simulations are generated with the Goddard Earth Observing System (GEOS-5) and confronted with Tb observations from the Soil Moisture Ocean Salinity (SMOS) mission. The MCMC algorithm suggests that the relative uncertainty of the RTM parameter estimates is typically less than 25% of the maximum a posteriori density (MAP) parameter value. Furthermore, the actual root-mean-square-differences in long-term Tb averages and standard deviations are found consistent with the respective estimated total simulation and observation error standard deviations of σ_m=3.1 K and σ_s=2.4 K. It is also shown that the MAP parameter values estimated through MCMC simulation are in close agreement with those obtained with Particle Swarm Optimization (PSO).

Keywords:

radiative transfer modeling, brightness temperature, Bayesian parameter estimation, uncertainty, Markov Chain Monte Carlo simulation, SMOS

1. Introduction

Uncertainties in radiative transfer modeling (RTM) affect the simulation of brightness temperatures (Tb) over land and the inversion of satellite-observed Tb to soil moisture retrievals. Quantification of these uncertainties is crucial to producing, validating and using passive microwave data, such as those obtained from the Soil Moisture Ocean Salinity (SMOS, Kerr et al. (2010)) and future Soil Moisture Active Passive (SMAP, Entekhabi et al. (2010)) missions. Yet, it is not particularly clear which RTM formulation and parameter values to use for large-scale applications.

In the context of forward Tb simulation, several studies have analyzed the effect of different RTM formulations for the microwave roughness length, vegetation parameterization and soil dielectric model (Drusch et al., 2009; de Rosnay et al., 2009). The impact of parameter values and dynamic land surface variables as input to large-scale forward Tb simulations was demonstrated by, e.g., De Lannoy et al. (2013) and Balsamo et al. (2006), respectively. Similarly, soil moisture retrievals based on Tb observations are affected by the RTM formulation and parameter values (Crosson et al., 2005; Panciera et al., 2009; Konings et al., 2011; Parinussa et al., 2011), as well as by the choice of background and auxiliary fields, such as soil temperature and vegetation characteristics (Kerr et al., 2012; O'Neill et al., 2012). Collectively, these studies suggest that RTMs exhibit significant uncertainty and that the precise magnitude and impact of this uncertainty on large-scale Tb simulations and soil moisture retrievals remain unclear.

Estimating the uncertainty of microwave RTM parameters is a major challenge, especially at larger spatial scales. Field experiments have provided RTM parameters values (de Rosnay et al., 2006; Grant et al., 2007; Panciera et al., 2009; Sabater et al., 2011), but mostly without uncertainty estimates. De Lannoy et al. (2013) derived global-scale RTM parameter values and ad hoc uncertainty estimates using SMOS observations and Particle Swarm Optimization (PSO, Kennedy and Eberhart (1995)). PSO is especially designed to find the optimal parameter values within a limited budget of function evaluations, but without recourse to estimating their underlying uncertainty.

In this paper, we introduce a (Bayesian) Markov chain Monte Carlo (MCMC) simulation method to estimate the posterior RTM parameter distribution. The DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm is used with parallel direction and snooker sampling from past states (Vrugt et al., 2008, 2009; Laloy and Vrugt, 2012), referred to as DREAM$_{\text{(ZS)}}$. Bayesian approaches such as DREAM$_{\text{(ZS)}}$ have many advantages over optimization methods such as PSO. The explicit treatment and analysis of uncertainty help to understand which parts of the RTM model are well resolved and which elements require further attention. Furthermore, a formal analysis of the residuals can be used to check the validity of our assumptions about the residual error distributions and to discern whether reliable parameter values have been derived.

The added value of obtaining posterior parameter distributions with Bayesian approaches, however, comes at an increased computational cost. Adequately sampling the posterior parameter distributions is too costly for global-scale operational applications that rely on evolving modeling systems in need of frequent re-calibrations, but can provide a valuable benchmark to verify results from simple parameter optimization algorithms, such as for example PSO.

The goals of the present paper are thus to infer RTM parameters and their posterior uncertainty using a Bayesian method, and to study the associated simulated Tb uncertainty. We are using the Goddard Earth Observing System (GEOS-5) modeling framework that will be used to generate the planned global SMAP Level 4 Surface and Root Zone Soil Moisture (L4_SM) data product through assimilation of SMAP Tb observations (Reichle et al., 2012). As in De Lannoy et al. (2013), we focus on optimizing time-invariant RTM-parameters by minimizing climatological differences between multi-angular, horizontally and vertically polarized Tb for morning and evening overpasses from SMOS observations and GEOS-5 simulations. The time-invariant optimized parameters will later be used in a data assimilation system (outside the scope of this paper), where state variables such as soil moisture and soil temperature will be updated in response to short-term variations in the observed Tb.

To summarize, in this paper we apply MCMC simulation using multi-angular SMOS Tb observations to (i) verify if the maximum a posteriori density (MAP) parameter values derived from a converged posterior distribution with DREAM$_{\mathrm{(ZS)}}$ can be approximated using PSO, (ii) obtain reliable parameter uncertainty estimates, and (iii) quantify the magnitude of parameter and other error sources in Tb simulations. The remainder of this paper is organized as follows. Section 2 summarizes the modeling system and the SMOS observations used in the present study. This is followed in section 3 by a description of the DREAM$_{\mathrm{(ZS)}}$ MCMC simulation method and PSO.

Section 3 also discusses several quantitative diagnostic metrics to analyze the simulated Tb uncertainty. Finally, this paper concludes in sections 4 and 5 with a discussion of the results and conclusions.

2. Observations and Model

2.1. SMOS Tb Data

Since its launch in November 2009, the SMOS mission provides global Tb data at a nominal spatial resolution of 43 km and with an equator overpass every 3 days. Here we use the multi-angular, full polarization Tb data from the period 1 July 2010 to 1 July 2012. Specifically, the data are extracted from the MIR_SCLF1C product, with processor version 504 for the years 2010 and 2011, and version 551 from January 2012 onwards. Our previous study presented in De Lannoy et al. (2013) discusses in detail the various steps involved in the processing of the SMOS data. Most importantly, the data are screened extensively using both product-based data quality information and model-based quality control rules. Furthermore, the data are spatially mapped onto a 36 km Equal-Area Scalable Earth Grid (EASE) and binned per incidence angle. Consistent with our previous study, only a subset of 6 incidence angles is used: θ=[32.5^o, 37.5^o, 42.5^o, 47.5^o, 52.5^o, and 57.5^o], where, for example, 32.5^o represents the average of all data with incidence angles between 32^o and 33^o.

For the purpose of estimating the microwave RTM parameters, long-term averages ($\boldsymbol{m}_o$) and standard deviations ($\boldsymbol{s}_o$) of the SMOS data are computed separately for each of the 6 incidence angles, 2 polarizations (horizontal H and vertical V), and 2 overpass times (ascending at 06:00h local time (LT), descending at 18:00h LT). This results in a total of 48 “observations” per grid cell: 24 for the long-term average Tb and 24 for the long-term Tb standard deviation. Section 3 provides more details.

2.2. GEOS-5 Tb Modeling

The modeling combines (i) land surface modeling with the Catchment land surface model (CLSM) and (ii) radiative transfer modeling with a tau-omega model to simulate long-term Tb averages and standard deviations. As in De Lannoy et al. (2013), the GEOS-5 CLSM (Koster et al., 2000) is set up on the 36 km EASE grid and spun up prior to the SMOS observation period. Surface meteorological forcing data at a $1/2^o \times 2/3^o$ spatial and hourly temporal resolution are taken from the Modern-Era Retrospective analysis for Research and Applications (MERRA, Rienecker et al. (2011)). The MERRA-precipitation is corrected with gauge-based precipitation from the National Oceanic and Atmospheric Administration (NOAA) Climate Prediction Center “Unified” (CPCU) product (Reichle, 2012). The model version is the same as that used for the MERRA-Land data product (Reichle et al., 2011), except for two changes that more closely align the model with the version that will ultimately be used for the SMAP L4_SM data product: (i) the surface soil moisture pertains to the top 5 cm surface layer (as opposed to the top 2 cm layer in MERRA-Land), and (ii) a preliminary version of updated soil parameters from a forthcoming version of GEOS-5 is used.

The vegetation parameterization in CLSM uses 8 default vegetation classes. For the RTM simulations, these classes are further refined into the 16 classes defined by the Moderate Resolution Imaging Spectroradiometer (500 m MOD12Q1 V004) International Geosphere-Biosphere Programme (IGBP) land cover classification (Loveland and Belward, 1997). Figure 1 shows the North American study domain which covers 9 of the 16 IGBP vegetation classes.

The soil moisture, soil temperature, vegetation water content, air temperature and climatological vegetation dynamics simulated with the prognostic CLSM are used as input to the diagnostic zero-order (tau-omega) microwave RTM to simulate L-band Tb. A short description of the RTM is given in Appendix. In essence, the Tb is determined by the surface soil temperature and attenuated by dynamic and static soil and vegetation characteristics. The key model parameters that impact the rough surface reflectivity h (Eq. A.3, Eq. A.4), the scattering albedo ω, and vegetation optical depth τ (Eq. A.6) will be estimated using the multi-angular SMOS observations (section 3), where h is a function of soil moisture and τ depends on the leaf area index (LAI).

3. Methods

3.1. Overview

Keeping up with our previous work (De Lannoy et al., 2013), the objective of the parameter estimation is to minimize the difference between long-term (climatological) averages and standard deviations for multiple types of SMOS-observed and GEOS-5-modeled Tb. We purposely do *not* minimize differences in the time domain as the goal of the present paper is to derive parameter estimates that result in the smallest possible bias in the long-term simulation of Tb. Short-term differences between Tb observations and simulations will be exploited in future studies using sequential data assimilation. We estimate a time-invariant multi-dimensional parameter set (hereafter referred to as $\boldsymbol{\alpha}$) that determines climatological features of the simulated Tb. The parameters are optimized locally, i.e., for each grid cell independently, and only for non-frozen land surface conditions as determined by the GEOS-5 modeling system.

Table 1 gives an overview of the parameters estimated in different experiments. All scenarios estimate the 5 most relevant RTM-parameters: h_{min}, $\Delta h \equiv h_{max} - h_{min}$, b_H, $\Delta b \equiv b_V - b_H$ and ω (according to the best scenario identified in De Lannoy et al. (2013)). Based on these *time-invariant* parameters, *time-variant* values of h, τ_H and τ_V are computed, using dynamic information about soil moisture for h (Eq. A.4) and LAI for τ (Eq. A.6). Time-averaged results for $< h >$ and $< \tau >$ are then presented, where $< \cdot >$ denotes the long-term time average. These RTM-parameters are estimated with either DREAM$_{\text{(ZS)}}$ or PSO, hereafter referred to as scenarios D and P, respectively. The DREAM$_{\text{(ZS)}}$ analysis is further expanded to also include the residual Tb error statistics σ_m and σ_s (scenario D$_\sigma$, discussed below). For each grid cell, we thus estimate 5 parameters for scenarios P and D, and 7 for D$_\sigma$.

To derive these parameters, we minimize per grid cell the *climatological*, or long-term, differences between 48 Tb observations and simulations. The 2×24 observations consist of long-term Tb averages and Tb standard deviations for the 24 combinations of 2 polarizations, 2 overpass times, and 6 incidence angles. The errors in these observations are assumed to be independent, that is, we neglect correlations in instrument errors and errors between H- and V-polarized observations at identical incidence angles. Similarly, the simulation errors are assumed to be independent, even though correlation is to be expected. Note that temporal correlations in the errors are of little concern because the observations are long-term averages and standard deviations, and not measurements in the time domain (Wöhling and Vrugt, 2011).

In keeping up with De Lannoy et al. (2013), the two years of historical SMOS data are divided into a calibration period (1 July 2011 - 1 July 2012) and an evaluation period (1 July 2010 - 1 July 2011). To ensure a meaningful calibration at each grid cell, we impose a minimum of 20 valid data points (N_i) per year to compute the long-term Tb average and standard deviation for a particular combination ($i = 1, \ldots, 24$) of incidence angle, polarization and overpass time. The requirement of $N_i \geq 20$ is used for the calculation of evaluation statistics as well.

3.2. Markov Chain Monte Carlo (MCMC) Sampling

The Bayesian framework allows deriving posterior probabilities of parameter estimates and model simulations, conditioned on errors in observations and simulations. The posterior probability distribution is computed by combining the observation likelihood $p(\boldsymbol{m}_o, \boldsymbol{s}_o|\boldsymbol{\alpha})$ with a prior distribution $p(\boldsymbol{\alpha})$:

$$p(\boldsymbol{\alpha}|\boldsymbol{m}_o, \boldsymbol{s}_o) = \frac{p(\boldsymbol{m}_o, \boldsymbol{s}_o|\boldsymbol{\alpha})p(\boldsymbol{\alpha})}{\int_{\boldsymbol{\alpha}} p(\boldsymbol{m}_o, \boldsymbol{s}_o|\boldsymbol{\alpha})d\boldsymbol{\alpha}} \tag{1}$$

The observations consist of long-term averages ($m_{i,o} \in \boldsymbol{m}_o$) and standard deviations ($s_{i,o} \in \boldsymbol{s}_o$) of Tb for 24 different combinations of incidence angles, polarizations and overpass times ($i = 1, \ldots, 24$). The denominator is a normalization factor and thus it suffices to maximize $p(\boldsymbol{m}_o, \boldsymbol{s}_o|\boldsymbol{\alpha})p(\boldsymbol{\alpha})$ to find the posterior distribution of $\boldsymbol{\alpha}$. In practice, it is difficult to solve this problem analytically and we therefore resort to MCMC simulation to generate a sample of the posterior target distribution.

In this paper, the differential evolution adaptive metropolis (DREAM$_{\text{(ZS)}}$, Vrugt et al. (2008); Laloy and Vrugt (2012)) algorithm with sampling from past states is used to efficiently explore the posterior parameter distribution. This algorithm adaptively updates the scale and orientation of the proposal distribution during sampling, and is specifically designed to rapidly explore multi-dimensional target distributions. In DREAM$_{\text{(ZS)}}$, multiple chains are running in parallel and the update of a chain is determined from an external sample of points that collectively summarizes the search history of all the individual chains. The log-likelihood of the current and proposed parameter values are compared using the Metropolis selection rule. If the proposal is accepted, the chain moves to this new point, otherwise the chain remains at its current position. Diminishing adaptation of the external archive of samples ensures convergence to the exact posterior distribution.

We assume a Gaussian prior for each of the individual parameters $\alpha_{0,k} \in \boldsymbol{\alpha}_0$. The mean and standard deviation of this multi-normal distribution $p(\boldsymbol{\alpha})$ are derived from literature values that yield reasonable Tb simulations compared to SMOS Tb and are summarized in Table 1. Note that these values were referenced as 'Lit2' in De Lannoy et al. (2013). The prior mean for each individual parameter is given by a vegetation-dependent value $\alpha_{0,k}$ and the standard deviation $\sigma_{\alpha_{0,k}}$ is defined by $\sigma^2_{\alpha_{0,k}} = (\alpha_{max,k} - \alpha_{min,k})^2/12$, using upper and lower bounds $[\alpha_{max,k}, \alpha_{min,k}]$.

The following log-likelihood function is used to minimize the differences in long-term Tb averages and standard deviations between observations $(m_{i,o}, s_{i,o})$

and simulations $(m_i(\boldsymbol{\alpha}), s_i(\boldsymbol{\alpha}))$:

$$
\begin{aligned}
L = \ln(p(\boldsymbol{m}_o, \boldsymbol{s}_o|\boldsymbol{\alpha})) = &\left. -\frac{24}{2}\ln(2\pi) - \frac{24}{2}\ln(\sigma_{i,m}^2) - \sum_{i=1}^{24}\frac{(m_{i,o} - m_i(\boldsymbol{\alpha}))^2}{2\sigma_{i,m}^2} \right\} L_{m,o} \\
&\left. -\frac{24}{2}\ln(2\pi) - \frac{24}{2}\ln(\sigma_{i,s}^2) - \sum_{i=1}^{24}\frac{(s_{i,o} - s_i(\boldsymbol{\alpha}))^2}{2\sigma_{i,s}^2} \right\} L_{s,o}
\end{aligned}
\tag{2}
$$

This formulation thus explicitly takes into consideration long-term biases in the Tb average ($L_{m,o}$ [-]) and the Tb variability ($L_{s,o}$ [-]) and is derived from a classical Gaussian likelihood function:

$$
\begin{aligned}
p(\boldsymbol{m}_o, \boldsymbol{s}_o|\boldsymbol{\alpha}) \;=\; & \prod_{i=1}^{24}\left[\frac{1}{\sqrt{2\pi\sigma_{i,m}^2}}\exp\left(-\frac{(m_{i,o} - m_i(\boldsymbol{\alpha}))^2}{2\sigma_{i,m}^2}\right)\right] \\
& \cdot\prod_{i=1}^{24}\left[\frac{1}{\sqrt{2\pi\sigma_{i,s}^2}}\exp\left(-\frac{(s_{i,o} - s_i(\boldsymbol{\alpha}))^2}{2\sigma_{i,s}^2}\right)\right]
\end{aligned}
\tag{3}
$$

where $\sigma_{i,m}$ and $\sigma_{i,s}$ denote the (ensemble) standard deviations of the residual differences between the observed and simulated values of the long-term Tb averages and standard deviations, respectively.

3.3. Particle Swarm Optimization (PSO)

The PSO algorithm (Kennedy and Eberhart, 1995) is a global search method that uses a dynamic swarm of particles to explore the parameter space. The best position of each individual particle (cognitive aspect) and of the entire swarm (social aspect) are used to guide the particles towards the optimal solution. The iterative swarm search is performed in several independent repetitions to account for sampling variability.

The fitness of each parameter combination in the swarm is measured by an integrated ‘cost’ or ‘objective function’ J [-] that measures the distances between the observed and simulated long-term Tb averages ($J_{m,o}$ [-]) and standard deviations ($J_{s,o}$ [-]). To make sure that the estimated parameter values honor the prior information (as used in DREAM$_{\text{(ZS)}}$), we also include a penalty term that quantifies deviations of the parameters from their expected values (J_α [-]). This results in the following definition of the objective function to be minimized:

$$
\begin{aligned}
J = {} & \left.\sum_{i=1}^{24} \frac{(m_{i,o} - m_i(\boldsymbol{\alpha}))^2}{2\sigma_{i,m}^2}\right\} J_{m,o} \\
& \left. + \sum_{i=1}^{24} \frac{(s_{i,o} - s_i(\boldsymbol{\alpha}))^2}{2\sigma_{i,s}^2} \right\} J_{s,o} \\
& \left. + \sum_{k=1}^{N_\alpha} \frac{(\alpha_{0,k} - \alpha_k)^2}{2\sigma_{\alpha_{0,k}}^2} \right\} J_\alpha
\end{aligned}
\tag{4}
$$

where N_α signifies the number of simultaneously estimated parameters. This formulation is essentially similar to the definition of the posterior density used in DREAM$_{\text{(ZS)}}$. The main difference is that PSO handles the prior information of the parameters explicitly as penalty term J_α in the objective function, whereas in DREAM$_{\text{(ZS)}}$, the prior parameter distribution is handled independently from the likelihood function by application of Bayes law. Both methods should thus find the same "best" parameter values.

3.4. Likelihood, Objective Function and Algorithm Settings

The design of the likelihood (L) or objective (J) function for DREAM$_{\text{(ZS)}}$ and PSO warrants further discussion. As discussed above, we sample the climatological, or long-term, Tb averages and standard deviations over multiple incidence angles, polarizations and overpass times (that is, 2 × 24 observations, $i = 1, \ldots, 24$) per location, rather than one observation at multiple time steps. The long-term Tb averages and standard deviations could also be interpreted as 'summary statistics' or 'signatures' of the system, and hence our approach has many elements in common with the diagnostic model evaluation procedure presented in Vrugt and Sadegh (2013).

The variables $\sigma_{i,m}$ and $\sigma_{i,s}$ in Eq. 2 and Eq. 4 measure the (ensemble) standard deviation of the residual differences between the observed and simulated long-term Tb averages and standard deviations, respectively, for each observation i. The residual errors are assumed to have a zero mean and include both SMOS observation and simulation errors, due to e.g. inaccurate soil moisture, temperature or vegetation characteristics. These $\sigma_{i,m}$ and $\sigma_{i,s}$ statistics trade-off errors in the long-term Tb averages against those of the long-term Tb standard deviations (as well as deviations from the prior parameter constraints). Since only one sample is available for each observation, it is impossible to estimate individual $\sigma_{i,m}$- and $\sigma_{i,s}$-values. Therefore, we define $\sigma_{i,m}$ and $\sigma_{i,s}$ as a combination of a homoscedastic term (σ_m, σ_s) and a tuning factor w_i to account for the robustness of the diagnosed long-term Tb averages and standard deviations, i.e. $\sigma^2_{i,m} = w_i\sigma^2_m$ and $\sigma^2_{i,s} = w_i\sigma^2_s$. The homoscedastic term is identical for all 24 observations and set to a default value of 1 K (De Lannoy et al., 2013), or alternatively we estimate σ_m and σ_s jointly with the RTM parameters (see section 3). The weights are given by $w_i = \frac{N}{N_i}$, where N_i denotes the number of data points in time that contribute to a particular long-term Tb average (or standard deviation), and N signifies the average number of time steps across all observations. These weights are typically close to 1 and assign somewhat more (less) weight to climatological differences that are based on more (fewer) individual data points in the original 1-year data time series.

A maximum of 12,000 log-likelihood function evaluations are performed with DREAM$_{\text{(ZS)}}$ using standard settings of the algorithmic variables. For PSO, we use the same algorithmic settings as De Lannoy et al. (2013), except a swarm size of 10 particles is used with a minimum of 10 and maximum of 100 iterations. The search is terminated if the reduction of the objective function is smaller than 1E-5 over the last 10 iterations. A total of 12 repetitions are performed, which results in a maximum of 12,000 function evaluations.

3.5. Posterior Parameter Distribution

The 'optimal' parameter values are defined as those with the maximal a posteriori density (MAP), i.e. with the largest value for L (Eq. 2, DREAM$_{\text{(ZS)}}$) or smallest value for J (Eq. 4, PSO). Note that these MAP values are not necessarily identical to the posterior ensemble mean of the distribution derived with of DREAM$_{\text{(ZS)}}$. For the DREAM$_{\text{(ZS)}}$ experiments, the last 25% of the MCMC chains (3,000 samples) are used to summarize parameter uncertainty by calculating the standard deviation of each individual parameter. To illustrate this in more detail for one grid cell, consider Fig. 2a, which depicts the marginal distributions of the RTM parameters. We define the uncertainty as the ensemble standard deviation $stdv[\alpha] \equiv \overline{\alpha - \overline{\alpha}}$ centralized around the ensemble mean $\overline{\alpha}$, not around the MAP parameter value α_{MAP}. The notation $\overline{\cdot}$ refers to the ensemble mean. Note that the standard deviation around the MAP estimate $stdv_{MAP}[.]$ can be found as a function of the centralized standard deviation $stdv[.]$, i.e. $stdv_{MAP}[.]^2 = stdv[.]^2 + \Delta_\alpha(\Delta_\alpha - stdv[.])$, where $\Delta_\alpha = \overline{\alpha} - \alpha_{MAP}$ is the difference between the ensemble mean and MAP parameter estimate. We found that, across the different experiments,

Δ_α is either small or Δ_α and $stdv[.]$ are of similar magnitude (not shown herein), so that $stdv_{MAP}[.] \sim stdv[.]$.

3.6. Convergence

'Convergence' can reflect accuracy (closeness to the actual optimum solution) or precision (reduction of the prior uncertainty). The following hypotheses will be verified to assess the convergence of the DREAM$_{\text{(ZS)}}$ algorithm: (i) the Tb performance (accuracy) with posterior parameter estimates should be better than with prior parameter guesses (section 3.7), (ii) the posterior parameter uncertainty (section 3.5) and the corresponding uncertainty in Tb simulations (section 3.7) should be reduced compared to their counterparts derived from the prior parameter uncertainty, and (iii) the potential scale reduction factor $\sqrt{R}$ by Gelman and Rubin (1992) should be near 1 to inspire confidence that the different MCMC chains have converged to the appropriate limiting distribution. The latter metric measures by which scale the posterior distribution will shrink as the number of MCMC iterations would go to infinity.

3.7. Tb Performance and Ensemble Verification

A number of measures are used to evaluate the long-term Tb simulations and their associated uncertainty. Fig. 2b illustrates some of the terms used in this evaluation. First, we assess the quality of the deterministic Tb simulations with the MAP parameter estimates, using the mean-square difference (MSD [K^2]) between the observed and simulated long-term Tb averages

(Eq. 5) and standard deviations (Eq. 6) across the 24 different observations:

$$MSD_m = \frac{1}{24}\sum_{i=1}^{24}(m_i(\boldsymbol{\alpha}_{MAP}) - m_{i,o})^2 \tag{5}$$

$$MSD_s = \frac{1}{24}\sum_{i=1}^{24}(s_i(\boldsymbol{\alpha}_{MAP}) - s_{i,o})^2 \tag{6}$$

If the modeling errors were solely due to uncertainties in the parameter values, these metrics should be very close to zero. In practice, however, the metrics will substantially deviate from zero and reflect residual errors that cannot be explained by parameter uncertainty. The 24 differences contributing to MSD_m are illustrated as Δ_{m_i} in Fig. 2b.

Secondly, we verify whether the spread in prior and posterior ensemble Tb simulations is in agreement with the misfit between modeled and observed values, in a mean-square sense. To this end, an ensemble of Tb simulations is generated by randomly drawing 20 samples from the prior and posterior parameter distributions. The misfit or skill is again defined using the mean-square difference (MSD [K^2]), but now for the ensemble means:

$$MSD_{\overline{m}} = \frac{1}{24}\sum_{i=1}^{24}\left(\overline{m_i(\boldsymbol{\alpha})} - m_{i,o}\right)^2 \tag{7}$$

$$MSD_{\overline{s}} = \frac{1}{24}\sum_{i=1}^{24}\left(\overline{s_i(\boldsymbol{\alpha})} - s_{i,o}\right)^2 \tag{8}$$

where $\overline{\cdot}$ denotes the ensemble mean. Fig. 2b illustrates the 24 differences contributing to $MSD_{\overline{m}}$ as $\Delta_{\overline{m_i}}$. If the uncertainties are well estimated and biases between observations and simulations are constrained during the calibration, the $MSD_{\overline{m}}$ and $MSD_{\overline{s}}$ metrics should match the total expected uncertainty ($MEnSp_m$, $MEnSp_s$), which is the sum of the Tb simulation spread due to parameter uncertainty ($EnSp_{i,m,par}$, $EnSp_{i,s,par}$) plus the residual Tb error variance ($\sigma^2_{i,m}$, $\sigma^2_{i,s}$):

$$MEnSp_m = \frac{1}{24}\sum_{i=1}^{24}\left[EnSp_{i,m,par} + \sigma^2_{i,m}\right] \tag{9}$$

$$MEnSp_s = \frac{1}{24}\sum_{i=1}^{24}\left[EnSp_{i,s,par} + \sigma^2_{i,s}\right] \tag{10}$$

where $\sigma^2_{i,m}$ and $\sigma^2_{i,s}$ are dominated by observation, input and structural error after the MAP parameters values have been found. The constituent terms $EnSp_{i,m,par}$ and $EnSp_{i,s,par}$ for each observation type i are given by:

$$EnSp_{i,m,par} = \overline{\left(m_i(\boldsymbol{\alpha}) - \overline{m_i(\boldsymbol{\alpha})}\right)^2} \tag{11}$$

$$EnSp_{i,s,par} = \overline{\left(s_i(\boldsymbol{\alpha}) - \overline{s_i(\boldsymbol{\alpha})}\right)^2} \tag{12}$$

An illustration of $EnSp_{i,m,par}$ is given in Fig. 2b. Again, if the uncertainties are well estimated, then the ratios $MSD_{\overline{m}}/MEnSp_m$ and $MSD_{\overline{s}}/MEnSp_s$ should be close to 1, or in other words: the "actual" ($MSD_{\overline{m}}$, $MSD_{\overline{s}}$) and "expected" ($MEnSp_m$, $MEnSp_s$) errors should be similar. These metrics are very similar to those used to verify the prescribed observation and simulation uncertainties in data assimilation systems (Reichle et al., 2002) and for ensemble forecast verification (De Lannoy et al., 2006). The only difference is that here, the mean values (i.e. the 'M', or mean, in MSD and $MEnSp$) are derived from multiple observations types ($i = 1, \ldots, 24$), whereas in the earlier studies the mean was calculated in the time domain.

4. Results

4.1. RTM-Parameters and Their Uncertainty

In this section, we analyze the MAP values of $< h >$, $< \tau >$ and ω, and their posterior uncertainty ($stdv[.]$). The DREAM$_{\text{(ZS)}}$ scenario D$_\sigma$ should be considered as benchmark in the following discussion, because of statistical rigor of the sampled posterior (will be further discussed below). Fig. 3 shows maps of the prior parameter values and the MAP estimates derived from scenario P (PSO), D and D$_\sigma$ (DREAM$_{\text{(ZS)}}$) (Table 1). The spatially averaged posterior parameter values are very similar for all 3 scenarios, with a microwave roughness $< h >$ around 0.75±0.5 [-], a nadir opacity $< \tau >$ of 0.26±0.15 [-] and a scattering albedo ω of 0.09±0.07 [-], where the values after the ± sign measure the spatial standard deviation and reflect the variability of the MAP parameters across the spatial domain. Note that these values should not be confused with uncertainty estimates. Compared to the prior values (Table 1 and 2), $< h >$ has generally increased for grassland, $< \tau >$ is smaller for forests and ω has increased for all vegetation classes except grassland (details per vegetation class not shown; these finding are similar to those of De Lannoy et al. (2013)). The spatial patterns for the 3 scenarios are also very similar. Moreover, Fig. 3 suggests that MAP values derived with the PSO algorithm closely match those of DREAM$_{\text{(ZS)}}$.

Fig. 4 shows the ensemble parameter uncertainty for scenarios D and D$_\sigma$. Maps with RTM parameter uncertainty estimates for PSO (obtained as in De Lannoy et al. (2013)) are not shown, because they are statistically invalid and significantly larger than those derived with DREAM$_{\text{(ZS)}}$. The relative uncertainties for scenario D are less than 10% of the MAP parameter value and substantially smaller than the spatial variability in the MAP values. For scenario D_σ, the relative uncertainties increase, with errors ranging up to 25% of the MAP values: for $< h >$, the spatially averaged uncertainty is 0.10 ± 0.08 [-], for $< \tau > 0.04 \pm 0.04$ [-] and for ω 0.02 ± 0.02 [-], respectively. The uncertainty in $< h >$ typically increases with more complex terrain and is smallest in the cropped region southwest of the Great Lakes. The uncertainty of $< \tau >$ is largest in the forested Appalachian mountains where the highest MAP values of $< \tau >$ are found. On the contrary, ω is best defined in this area and uncertainties in ω increase in the Western dry mountain ranges. The $< h >$-values are more uncertain where either the uncertainty in ω (Fig. 4e) or $< \tau >$ (Fig. 4f) is larger.

In summary, both $\text{DREAM}_{(\text{ZS})}$ scenarios D and D_σ provide MAP parameter values that are very similar and in close agreement with the PSO estimates. The $\text{DREAM}_{(\text{ZS})}$ derived posterior parameters appear well defined with relative uncertainties that are less than 25% of the MAP values. It will be shown below that the uncertainty estimates of scenario D_σ are consistent with the sample RMSD between long-term Tb observations and simulations.

4.2. Residual Tb Error Standard Deviation Estimation

To analyze the effect of σ_m and σ_s in more detail, Table 2 summarizes the MAP parameter values and their associated uncertainties averaged over the entire study domain. In addition, Fig. 5 depicts the results for different vegetation classes. As discussed above, scenarios D and D_σ return similar MAP RTM-parameter values, but when σ_m and σ_s are simultaneously estimated, the posterior RTM-parameter uncertainty increases about 2 - 3 times. The domain-averaged values for scenario D_σ are $\sigma_m = 3.1$ K and $\sigma_s = 2.4$ K

(Table 2), whereas scenario D uses default values of these variables of 1 K.

The value of σ_m and its posterior uncertainty are largest in cropped regions (Fig. 5g) where residual Tb errors are dominated by less skillful model simulations. This is to be expected because irrigation is not simulated and the climatological LAI estimates do not account for interannual crop rotations. The parameters can not compensate for these errors, and the default values of $\sigma_m = \sigma_s = 1$ K make scenarios D and P vulnerable to suboptimal solutions. For example, the relative large differences between D and D_σ for σ_m and σ_s over cropland areas increases the differences in the MAP values of ω. For forests, $\sigma_s = 1$ K appears to be a good estimate (Fig. 5i) because the variability in Tb is expected to be low due to vegetation attenuation. Both the MAP values and uncertainties for σ_m are always larger than those derived for σ_s. One of the reasons for the higher σ_m are the opposite signs in the biases for the long-term averages of ascending and descending Tb, which cannot be mitigated with time-invariant RTM-parameters. These biases are due to sensor error and modeled temperature errors as discussed in De Lannoy et al. (2013). In a separate exercise (not shown herein), we verified that the σ-values absorb biases in geophysical fields: by re-scaling the soil moisture both the $RMSD$ (see below) and σ-values are jointly reduced.

For the simulations with prior parameters, we also calculated (i.e. not optimized) σ_m and σ_s as 7.5 K and 4.8 K, respectively (Table 2). Unlike the MAP σ_m- and σ_s-values, these prior residual σ-values are dominated by simulation error due to suboptimal parameter values.

4.3. MAP Tb Performance

Fig. 6 shows the misfit between observed and MAP simulated long-term Tb averages and standard deviations ($RMSD_m$, $RMSD_s$, square-root of Eq. 5 and 6) across the 24 observations for the calibration and evaluation period, averaged per vegetation class. The performance skill is very similar for scenarios P, D and D_σ, which reflects that the three scenarios generate similar parameter estimates. The $RMSD_m$ ranges between 2 and 4.5 K during the calibration (Fig. 6a) and increases up to 8 K for cropland in the evaluation period (Fig. 6c). The $RMSD_s$ ranges between 1 and 3 K during calibration (Fig. 6b) and reaches values of 5 K for cropland in the evaluation year (Fig. 6d). Cropland has the highest errors, because of known simulation errors (see above). Note also that the $RMSD_m$ and $RMSD_s$ values of scenario D_σ during the calibration period (Fig. 6a-b) show the same pattern as σ_m and σ_s in Fig. 5g and 5i. The increased errors in the evaluation period suggest that the calibration could benefit from climatological observations based on longer data records to better estimate the parameter values.

4.4. Ensemble Tb Performance

For $\text{DREAM}_{(\text{ZS})}$, we analyze the balance between the actual Tb misfit and the expected uncertainty (ensemble variance) in the ensemble Tb simulations (20 members, as opposed to single deterministic MAP simulations above). The results are presented in Table 2 and Fig. 7. Table 2 shows the skill of the ensemble mean Tb simulations $\overline{m_i(\boldsymbol{\alpha})}$ and $\overline{s_i(\boldsymbol{\alpha})}$ for the calibration period in terms of $RMSD_{\overline{m}}$ and $RMSD_{\overline{s}}$, i.e. the square-root of Eqs. 7 and 8. These values are very similar to the results for the MAP simulations (section 4.3). For both scenarios D and D_σ, the $RMSD_{\overline{m}}$ and $RMSD_{\overline{s}}$ are respectively

3 K and 2.5 K, which is less than half of the actual misfit when the prior parameters are used.

Table 2 also lists the square-root of the combined mean simulation and observation spread or expected uncertainty, i.e. $RMEnSp_m$ and $RMEnSp_s$ (square-root of Eqs. 9 and 10), along with the constituent terms ($RMEnSp_{m,par}$, $RMEnSp_{s,par}$, σ_m and σ_s). Generally, the uncertainty associated with the parameter values is much smaller than the uncertainty related to other factors, that is, $RMEnSp_{par} < \sigma$, which is valid both when using prior and posterior parameter distributions. Moreover, after calibration both the σ- and $RMEnSp_{par}$-values are significantly reduced compared to their prior values.

If the uncertainty estimates are consistent, $RMSD_{\overline{m}} \sim RMEnSp_m$ and $RMSD_{\overline{s}} \sim RMEnSp_s$, i.e. there should be a balance between the actual and expected errors (section 3.7). The domain-averaged $RMSD_{\overline{m}}/RMEnSp_m$ is 2.7 for scenario D and 1.0 for scenario D_σ. Similarly, the domain-averaged $RMSD_{\overline{s}}/RMEnSp_s$ is 2.5 for scenario D and 1.0 for scenario D_σ. Optimal results are thus only found after including an estimation of σ_m and σ_s in scenario D_σ. Note that for the evaluation period (not shown), the ratios always exceed 1, because of an increased $RMSD_{\overline{m}}$ and $RMSD_{\overline{s}}$.

Fig. 7 shows how the ensemble spread is consistent with misfits between observations and simulations for scenario D_σ. Specifically, Fig. 7a shows the SMOS observed $m_{i,o}$ and the GEOS-5 simulated $\overline{m_i(\boldsymbol{\alpha})}$ for ascending, H-polarized Tb at 6 angles for scenarios D and D_σ, averaged over the entire study domain. Fig. 7b shows the same for V-polarized Tb, and Figs. 7c and d provide this information for the long-term Tb standard deviations.

Also shown is the total ensemble simulation and observation uncertainty for each observation type, presented as error bars around the ensemble mean Tb simulations for illustration.

The error-bars for scenario D_σ fully envelop the observations, whereas this is not the case for scenario D. Fig. 7 also explains the nature of the residual misfit. Except for the 57.5°-angle, the ascending ensemble mean simulations $\overline{m_i(\boldsymbol{\alpha})}$ consistently underestimate the SMOS-observed $m_{i,o}$ for H-polarization and randomly deviate from the SMOS-observed $m_{i,o}$ at V-polarization. In contrast, the descending simulations $\overline{m_i(\boldsymbol{\alpha})}$ slightly overestimate the SMOS-observed $m_{i,o}$ at H-polarization (see De Lannoy et al. (2013)). The SMOS-observed $s_{i,o}$ is always larger than the simulated $\overline{s_i(\boldsymbol{\alpha})}$. This is probably dominated by observation noise, but could also be attributed to an underestimated variability in the Tb simulations. For example, an increase in the RTM-parameter h not only compensates for a cold bias but simultaneously also reduces the Tb variability. Fig. 7 clearly illustrates why the uncertainty estimates obtained from scenario D_σ are superior.

4.5. Convergence and Computational Cost

The effectiveness of the posterior parameter sampling is measured by the convergence of the algorithms. Table 2 confirms that both the posterior uncertainties in the parameter estimates ($stdv[.]$) and the misfit between the simulations and observations ($RMSD$) of the long-term Tb averages and standard deviations are greatly reduced compared to the results with the prior parameter distribution. Another measure for convergence is the scale reduction factor, or $\sqrt{R}$-statistic by Gelman and Rubin (1992). Values close to 1 are preferred, and suggest that the MCMC sampler has converged to a limiting distribution. Fig. 8 shows the evolution of the convergence diagnostic $\sqrt{R}$ for both DREAM$_{(\mathrm{ZS})}$ scenarios. The $\sqrt{R}$ is averaged over all estimated parameters and across the study domain, since no obvious differences in $\sqrt{R}$ are found between the different vegetation classes (not shown). Initially, the values of $\sqrt{R}$ exhibit a lot of variation (due to random initial sample) before they settle down and reach values close to 1.

Finally, we report that the derivation of the posterior distributions requires approximately 225 seconds for a single grid cell using DREAM$_{(\mathrm{ZS})}$. For global applications that involve $10^5 - 10^6$ grid cells, posterior distribution exploration may be too costly. Yet, if we target the MAP value only, PSO or DREAM$_{(\mathrm{ZS})}$ are both viable options.

5. Conclusions

Accurate estimates of microwave RTM parameters for large-scale L-band applications are difficult to obtain. The available parameter estimates are generally based on small-scale field experiments and often come without any estimate of posterior uncertainty. This complicates radiative transfer modeling for both the forward simulation of L-band Tb over land and the retrieval of soil moisture based on Tb observations. This paper expands earlier research reported in De Lannoy et al. (2013) to derive time-invariant RTM-parameters using observations of the long-term average Tb and the long-term Tb standard deviation obtained from SMOS data. The overall objective is to optimize GEOS-5 Tb simulations prior to sequential assimilation of SMOS or SMAP Tb data, such as planned for the SMAP L4_SM product (Reichle et al., 2012) and to examine the uncertainties involved in the optimization.

Per grid cell, 48 observations of the long-term Tb averages and standard deviations were constructed for 24 different combinations of 6 incidence angles, 2 polarizations and 2 overpass times. The differences with their respective long-term GEOS-5 simulations are minimized (as opposed to minimizing differences between Tb observations and simulations in the time domain) and used along with the prior parameter information to derive posterior parameter estimates.

In the present paper, the full posterior distribution of RTM-parameters is derived using MCMC simulation with the DREAM$_{\text{(ZS)}}$ algorithm. To our knowledge, this is the first large-scale application of the DREAM$_{\text{(ZS)}}$ algorithm for the estimation of RTM-parameters and their underlying uncertainty. The results serve as a benchmark to verify the results from simpler parameter optimization algorithms, such as for example PSO. Simple algorithms are desirable for global-scale operational applications that rely on evolving modeling systems in need of frequent re-calibrations.

First, we verified that the MAP RTM-parameter values derived from converged posterior distributions with DREAM$_{\text{(ZS)}}$ can be approximated by a simpler optimization algorithm (PSO), which corroborates our earlier research (De Lannoy et al., 2013). Secondly, we obtained reliable parameter uncertainty estimates with DREAM$_{\text{(ZS)}}$, which are impossible to estimate with PSO. The relative parameter uncertainties are generally less than 25% of the MAP value for $< h >$, $< \tau >$ and ω, when including the residual (observation and simulation) error statistics (σ_m, σ_s) of the long-term Tb averages and standard deviations in the estimation.

The third objective of this paper was to quantify the importance of parameter and other errors on Tb simulations. The uncertainty associated with the parameter values only contributes a small part to the total Tb uncertainty. Most of the discrepancy between Tb simulations and observations is covered by residual Tb errors, with MAP estimates of the standard deviations σ_m and σ_s (assumed homoscedastic) around 3.1 K and 2.4 K, respectively. The prior estimate of 1 K was thus too low, except for σ_s over forests which exhibit limited Tb variability due to vegetation attenuation. The largest σ_m-values are found in cropped regions where the RMSD between Tb simulations and observations is also highest, due to observation errors and errors in geophysical fields (e.g. soil moisture and temperature) that constitute important inputs to the Tb simulations.

The expected Tb error, i.e. the total of the MAP residual Tb error variance estimates (σ_m^2, σ_s^2) and the Tb spread introduced by the posterior parameter uncertainties ($EnSp_{i,m,par}$, $EnSp_{i,s,par}$), is found to be consistent with the actual RMSD of 3 and 2.5 K for the long-term posterior Tb averages and standard deviations. In other words, the joint estimation of RTM-parameters, σ_m and σ_s with DREAM$_{\text{(ZS)}}$ results in a balance between actual and expected errors in Tb simulations, and in statistically adequate parameter values and uncertainty estimates.

In summary, the Bayesian inference of the posterior distribution of the RTM-parameters ensures reliable Tb simulations with GEOS-5. Furthermore, the DREAM$_{\text{(ZS)}}$ algorithm also reveals the importance of observation error and simulation error that cannot be explained by the RTM parameters. These error sources can be addressed using model refinement and assimilation of satellite-observed Tb data.

Acknowledgement

The authors thank Yann Kerr and Ali Mahmoodi for their help with the SMOS data.

References

Balsamo, G., J, F. M., Bélair, S., Deblonde, G., 2006. A global root-zone soil moisture analysis using simulated L-band brightness temperature in preparation for the hydros satellite mission. Journal of Hydrometeorology 7, 1126–1146.

CESBIO, IPSL, INRA, Reading University, Tor Vergata University, 2011. SMOS level 2 processor for soil moisture. Tech. rep., CEBIO, IPLS-Service d'Aeronomie, INRA-EPHYSE, Reading University, Tor Vergata University.

Choudhury, B. J., Schmugge, T. J., Chang, A., Newton, R. W., 1979. Effect of surface roughness on the microwave emission from soils. Journal of Geophysical Research 84 (C9), 5699–5706.

Crosson, W., Limaye, A., Laymon, C., 2005. Parameter sensitivity of soil moisture retrievals from airborne L-band radiometer measurements in SMEX02. IEEE Transactions on Geoscience and Remote Sensing 43, 1517–1528.

De Lannoy, G., Reichle, R., Pauwels, V., 2013. 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, http://dx.doi.org/10.1175/JHM-D-12-092.1.

De Lannoy, G. J. M., Houser, P. R., Pauwels, V. R. N., Verhoest, N. E. C., 2006. Assessment of model uncertainty for soil moisture through ensemble verification. Journal of Geophysical Research 111 (D10), D10101.1–18, doi:10.1029/2005JD006367.

de Rosnay, P., Calvet, J., Kerr, Y., Wigneron, J., Lematre, F., Escorihuela, M., Muoz-Sabater, J., Saleh, K., Barri, J., Bouhours, G., Coret, L., Cherel, G., Dedieu, G., Durbe, R., Fritz, N., Froissard, F., Hoedjes, J., Kruszewski, A., Lavenu, F., Suquia, D., Waldteufel, P., 2006. SMOSREX: A long term field campaign experiment for soil moisture and land surface processes remote sensing. Remote Sensing of Environment 102, 377–389, doi:10.1016/j.rse.2006.02.021.

de Rosnay, P., Drusch, M., Boone, A., Balsamo, G., Decharme, B., Harris, P., Kerr, Y., Pellarin, T., Polcher, J., Wigneron, J., 2009. AMMA land surface model intercomparison experiment coupled to the Community Microwave Emission Model: ALMIP-MEM. Journal of Geophysical Research 114 (D05108), 1–18, doi:10.1029/2008JD010724.

Dobson, M., Ulaby, F., Hallikainen, M., El-Rayes, M. A., 1985. Microwace dielectric behavior of wet soil - part II: Dielectric mixing models. IEEE Transactions on Geoscience and Remote Sensing 23, 35–46.

Drusch, M., Holmes, T., de Rosnay, P., Balsamo, G., 2009. 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, 213–226.

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., Van Zyl, J., 2010. The Soil Moisture Active and Passive (SMAP) mission. Proceedings of the IEEE 98 (5), 704–716.

Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences. Stat. Sci. 7, 457–472.

Grant, J., Wigneron, J., Van de Griend, A. A., Kruszewski, A., Sobjaerg, S. S., Skou, N., 2007. A field experiment on microwave forest radiometry: L-band signal behaviour for varying conditions of surface wetness. IEEE Transactions on Geoscience and Remote Sensing 109, 10–19.

Jackson, T., Schmugge, T. J., 1991. Vegetation effects on the microwave emission of soils. Remote Sensing of Environment 36, 203–212.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. In: Proc. Int. Conf. Neur. Netw. Piscataway, NJ.

Kerr, Y., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N., Gruhier, C., Juglea, S., Drinkwater, M., Hahne, A., Martin-Neira, M., Mecklenburg, S., 2010. The SMOS mission: New tool for monitoring key elements of the global water cycle. Proceedings of the IEEE 98 (5), 666–687.

Kerr, Y. H., Njoku, E. G., 1990. A semiempirical model for interpreting microwave emission from semiarid land surfaces as seen from space. IEEE Transactions on Geoscience and Remote Sensing 28 (3), 384–393.

Kerr, Y. H., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., Al Bitar, A., Cabot, F., Gruhier, C., Juglea, S. E., Leroux, D., Mialon, A., Delwart, S., 2012. The SMOS soil moisture retrieval algorithm. IEEE Transactions on Geoscience and Remote Sensing 50 (5), 1384–1403.

Konings, A., Entekhabi, E., Chan, S. K., Njoku, E. G., 2011. The effect of rtm-parameter on satellite-based soil moisture retrievals. IEEE Transactions on Geoscience and Remote Sensing 49, 2686–2698.

Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M., Kumar, P., 2000. A catchment-based approach to modeling land surface processes in a general circulation model 1. model structure. Journal of Geophysical Research 105 (D20), 24809–24822.

Laloy, E., Vrugt, J., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM_(ZS) and high-performance computing. Water Resources Research 48 (W01526), doi:10.1029/2011WR010608.

Loveland, T. R., Belward, A. S., 1997. The IGBP-DIS global 1-km land cover data set, DIScover: first results. International Journal of Remote Sensing 65, 1031–1031.

Mironov, V., Dobson, M., Kaupp, V., Komarov, S., Kleshchenko, V., 2004. Generalized refractive mixing dielectric model for moist soil. IEEE Transactions on Geoscience and Remote Sensing 42, 773–785.

O'Neill, P., Njoku, E., Jackson, T., Chan, S., Bindlish, R., 2012. SMAP Algorithm Theoretical Basis Document: L2 & L3 soil moisture (passive) products. Tech. Rep. SMAP Project, JPL D-66481, Jet Propulsion Laboratory, Pasadena, CA.

Panciera, R., Walker, J. P., Merlin, O., 2009. Improved understanding of soil surface roughness parameterization for L-band passive microwave soil moisture retrieval. IEEE Transactions on Geoscience and Remote Sensing 6 (4), 625–629.

Parinussa, R. M., Meesters, A. G., Liu, Y. Y., Dorigo, W., Wagner, W., de Jeu, R. A., 2011. Error estimates for near-real-time satellite soil moisture as derived from the land parameter retrieval model. IEEE Transactions on Geoscience and Remote Sensing 8 (4), 779–783.

Pellarin, T., Wigneron, J., Calvet, J., Berger, M., Douville, H., Ferrazzoli, P., Kerr, Y. H., Lopez-Baeza, E., Pulliainen, J., Simmods, L. P., Waldteufel, P., 2003. Two-year global simulation of L-band brightness temperatures over land. IEEE Transactions on Geoscience and Remote Sensing 42 (9), 2135–2139.

Reichle, R., Crow, W., Koster, R., Kimball, J., De Lannoy, G., 2012. SMAP Algorithm Theoretical Basis Document: L4 surface and root zone soil moisture product. Tech. Rep. SMAP Project, JPL D-66483, Jet Propulsion Laboratory, Pasadena, CA.

Reichle, R. H., 2012. The MERRA-land data product (version 1.2), GMAO office note No. 3. Tech. rep., NASA Global Modeling and Assimilation Office, available at `http://gmao.gsfc.nasa.gov/pubs/office_notes/`.

Reichle, R. H., Koster, R. D., De Lannoy, G. J. M., Forman, B. A., Liu, Q., Mahanama, S. P. P., Toure, A., 2011. Assessment and enhancement of MERRA land surface hydrology estimates. Journal of Climate 24, 6322–6338.

Reichle, R. H., Walker, J. P., Koster, R. D., Houser, P. R., 2002. Extended vs. ensemble kalman filtering for land data assimilation. Journal of Hydrometeorology 3, 728–740.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M., Woollen, J., 2011. MERRA - NASA's modern-era retrospective analysis for research and applications. Journal of Climate 24 (14), 3624–3648, doi:10.1175/JCLI-D-11-00015.

Sabater, J. M., de Rosnay, P., Balsamo, G., 2011. Sensitivity of L-band NWP forward modelling to soil roughness. International Journal of Remote Sensing iFirst, 1–14, doi:10.1080/01431161.2010.507260.

Vrugt, J., Sadegh, M., 2013. Towards diagnostic model calibration and evaluation: Approximate Bayesian Computation. Water Resources Research 49, doi: 10.1002/wrcr.20354.

Vrugt, J., ter Braak, C., Clark, M., Hyman, J., Robinson, B., 2008. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. Water Resources Research 44 (W00B09), doi:10.1029/2007WR006720.

Vrugt, J. A., ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., Higdon, D., 2009. Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. International Journal of Nonlinear Sciences & Numerical Simulation 10 (3), 271–288.

Wang, J. R., Choudhury, B. J., 1981. Remote sensing of soil moisture content over bare field at 1.4 GHz frequency. Journal of Geophysical Research 86, 5277–5282.

Wang, J. R., Schmugge, T. J., 1980. An empirical model for the complex dielectric permittivity of soils as a function of water content. IEEE Transactions on Geoscience and Remote Sensing GE-18 (4), 288–295.

Wöhling, T., Vrugt, J. A., 2011. Multiresponse multilayer vadose zone model calibration using Markov chain Monte Carlo simulation and field water retention data. Water Resources Research 47, W04510.1–19.

Appendix A. Radiative Transfer Model

A diagnostic zero-order (tau-omega) microwave RTM is used to simulate L-band Tb at the top of the atmosphere ($\mathrm{Tb}_{TOA,p}$ [K]). The $\mathrm{Tb}_{TOA,p}$ at polarization $p = [H, V]$ (horizontal or vertical) is a combination of (i) soil emission, possibly attenuated by vegetation, (ii) vegetation emission, possibly reflected by the soil, and (iii) atmospheric effects:

$$\begin{aligned}
\mathrm{Tb}_{TOV,p} &= \mathrm{T}_s(1-r_p)A_p + \mathrm{T}_c(1-\omega_p)(1-A_p)(1+r_pA_p) \\
&\quad +\mathrm{Tb}_{ad,p}r_pA_p^2 \qquad \text{(A.1)} \\
\mathrm{Tb}_{TOA,p} &= \mathrm{Tb}_{au,p} + \exp(-\tau_{atm,p})\mathrm{Tb}_{TOV,p} \qquad \text{(A.2)}
\end{aligned}$$

where $\mathrm{Tb}_{TOV,p}$ [K] is the top of vegetation Tb, T_s [K] is the surface soil temperature, T_c [K] is the canopy temperature (assumed equal to T_s), $\mathrm{Tb}_{ad,p}$ [K] and $\mathrm{Tb}_{au,p}$ [K] are the downward and upward atmospheric radiation, A_p [-] is the vegetation attenuation, $\exp(-\tau_{atm,p})$ [-] is the atmospheric attenuation, $\tau_{atm,p}$ [-] is the atmospheric optical depth, r_p [-] is the rough surface reflectivity, and ω_p [-] is the scattering albedo. The atmospheric contributions ($\mathrm{Tb}_{ad,p}$, $\mathrm{Tb}_{au,p}$ and $\exp(-\tau_{atm,p})$) are described by Pellarin et al. (2003). The rough surface reflectivity r_p [-] is derived from the smooth surface reflectivity R_p [-] following (Choudhury et al., 1979; Wang and Choudhury, 1981):

$$r_p = (Q \;\; R_q + (1-Q)R_p)\exp(-h)\cos^{Nr_p}(\theta) \qquad \text{(A.3)}$$

where Q [-] is the polarization mixing ratio and typically set to 0 for L-band (Kerr and Njoku, 1990), θ [rad] is the incidence angle, h [-] is the roughness parameter accounting for dielectric properties that vary at the sub-wavelength scale, Nr_p [-] is the angular dependence, and $q = V$ for $p = H$

and vice versa. The smooth surface reflectivity R_p [-] is given by the Fresnel equations as a function of the dielectric constant, which itself depends on soil moisture, temperature, texture, incidence angle and wavelength. We select the Wang and Schmugge (1980) soil dielectric mixing model for this study. The results with this model are similar to what is obtained with the Mironov et al. (2004) model, and both are in a better agreement with the SMOS data than the Dobson et al. (1985) model. We include the dependence of h on soil moisture (SM [m^3.m^{-3}]) through a stepwise linear expression (adapted from the proposed SMOS soil moisture retrieval algorithm (CESBIO et al., 2011; Kerr et al., 2012)):

$$h = \begin{cases} h_{max} & \text{if } SM <= wt \\ h_{max} + \frac{h_{min}-h_{max}}{poros-wt}\,(SM - wt) & \text{if } wt < SM <= poros \end{cases} \tag{A.4}$$

where $poros$ [m^3.m^{-3}] and wt [m^3.m^{-3}] are the porosity and transition soil moisture, respectively. The latter is modeled as $wt = 0.48.wp + 0.165$ (Wang and Schmugge (1980)) where wp [m^3.m^{-3}] is the wilting point.

The vegetation attenuation A_p [-] is based on the Jackson and Schmugge (1991) vegetation opacity model:

$$A_p = \exp(-\frac{\tau_p}{\cos\theta}), \text{with} \tag{A.5}$$

$$\tau_p = b_p\, VWC = b_p\, LEWT\; LAI \tag{A.6}$$

where τ_p [-] is the nadir vegetation opacity, which is a function of a vegetation structure parameter b_p [-] and the vegetation water content (VWC) [kg.m^{-2}]. The latter is modeled here as the product of LAI [m^2.m^{-2}] and the leaf equivalent water thickness ($LEWT$) [kg.m^{-2}].

List of Tables

List of Figures

Table 1: Parameters selected for calibration in different scenarios (P, D, D_σ) with an indication of the allowed parameter range and the prior estimate for each IGBP vegetation class within the study domain. $\Delta h \equiv h_{max} - h_{min}$, $\Delta b \equiv b_V - b_H$.

		h_{min} [-]	$\Delta h[-]$	ω [-]	b_H [-]	Δb [-]	σ_m [K]	σ_s [K]
	P, D	X	X	X	X	X	-	-
	D_σ	X	X	X	X	X	X	X
	α_{min}	0	0	0	0	-0.15	1E-5	1E-5
	α_{max}	2.0	1.0	0.3	0.7	0.15	60	40
ENF	Evergreen Needleleaf Forest	1.2	0	0.05	0.33	0	1	1
DBF	Deciduous Broadleaf Forest	1	0	0.05	0.33	0	1	1
MXF	Mixed Forest	1.3	0	0.05	0.33	0	1	1
CSH	Closed Shrublands	0.7	0	0.05	0.3	0	1	1
OSH	Open Shrublands	0.7	0	0.05	0.3	0	1	1
WSV	Woody Savannas	0.7	0	0.05	0.3	0	1	1
GRS	Grasslands	0.1	0	0.05	0.2	0	1	1
CRP	Croplands	0.5	0	0.05	0.15	0	1	1
CRN	Crop and Natural Vegetation	0.7	0	0.05	0.15	0	1	1

Table 2: Domain-averaged parameters values and their uncertainty $stdv[.]$ for the prior distributions and the posterior distributions obtained with scenarios P, D and D_σ. The bottom half of the table shows ensemble Tb prediction statistics (square-root of Eq. 7-8, 9-10 and 11-12), averaged across 24 long-term Tb observations and calculated for the calibration period. Only for the prior parameters, σ_m and σ_s are calculated assuming (a) $RMEnSp_m = RMSD_{\overline{m}}$ and $RMEnSp_s = RMSD_{\overline{s}}$, and (b) $\sigma_m = \sqrt{RMSD^2_{\overline{m}} - RMEnSp^2_{m,par}}$ and $\sigma_s = \sqrt{RMSD^2_{\overline{s}} - RMEnSp^2_{s,par}}$.

	Prior	P	D	D_σ
$< h >$ [-]	0.59	0.74	0.75	0.77
$< \tau >$ [-]	0.35	0.26	0.26	0.25
ω [-]	0.05	0.09	0.09	0.09
σ_m [K]	7.45^b	1.00	1.00	3.08
σ_s [K]	4.78^b	1.00	1.00	2.39
$stdv[< h >]$ [-]	0.63	-	0.04	0.10
$stdv[< \tau >]$ [-]	0.27	-	0.02	0.04
$stdv[\omega]$ [-]	0.09	-	0.01	0.02
$stdv[\sigma_m]$ [K]	-	-	-	0.71
$stdv[\sigma_s]$ [K]	-	-	-	0.53
$RMSD_{\overline{m}}$ [K]	7.63	-	2.77	3.02
$RMSD_{\overline{s}}$ [K]	5.04	-	2.53	2.54
$RMEnSp_m$ [K]	7.63^a	-	1.04	3.24
$RMEnSp_s$ [K]	5.04^a	-	1.01	2.45
$RMEnSp_{m,par}$ [K]	1.65	-	0.28	0.92
$RMEnSp_{s,par}$ [K]	1.57	-	0.14	0.39

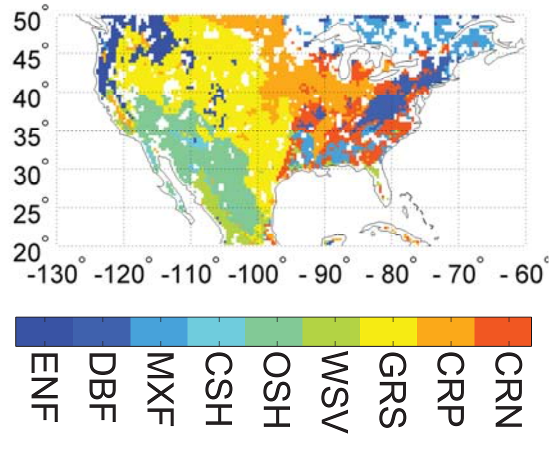

Figure 1: Study domain with indication of the dominant IGBP vegetation classes.

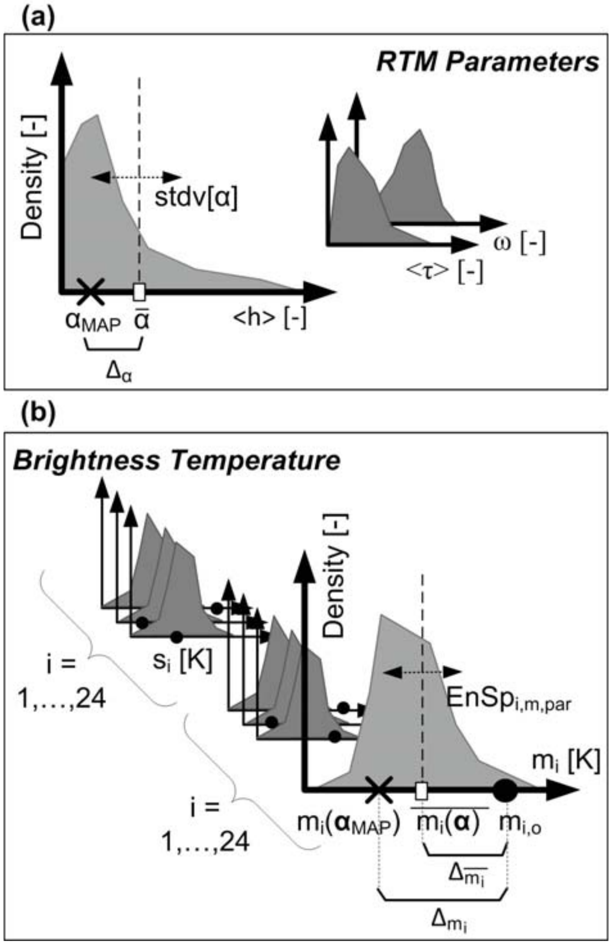

Figure 2: Illustration of marginal distributions for (a) RTM-parameters and (b) Tb simulations at a single grid cell. Crosses (×) indicate the MAP estimates, the vertical dashed lines and white box indicate the ensemble mean posterior estimate, and horizontal dotted arrows indicate one standard deviation uncertainty around the ensemble mean. The performance of the Tb simulations is quantified by comparing either the MAP ($m_i(\boldsymbol{\alpha}_{MAP})$, $s_i(\boldsymbol{\alpha}_{MAP})$) or the ensemble mean ($\overline{m_i(\boldsymbol{\alpha})}$, $\overline{s_i(\boldsymbol{\alpha})}$) simulations against (black dots) 24 observed values ($m_{i,o}, s_{i,o}$) with $i = 1, \ldots, 24$. The differences Δ_{m_i} and $\Delta_{\overline{m}_i}$ contribute to MSD_m (Eq. 5) and $MSD_{\overline{m}}$ (Eq. 7), respectively.

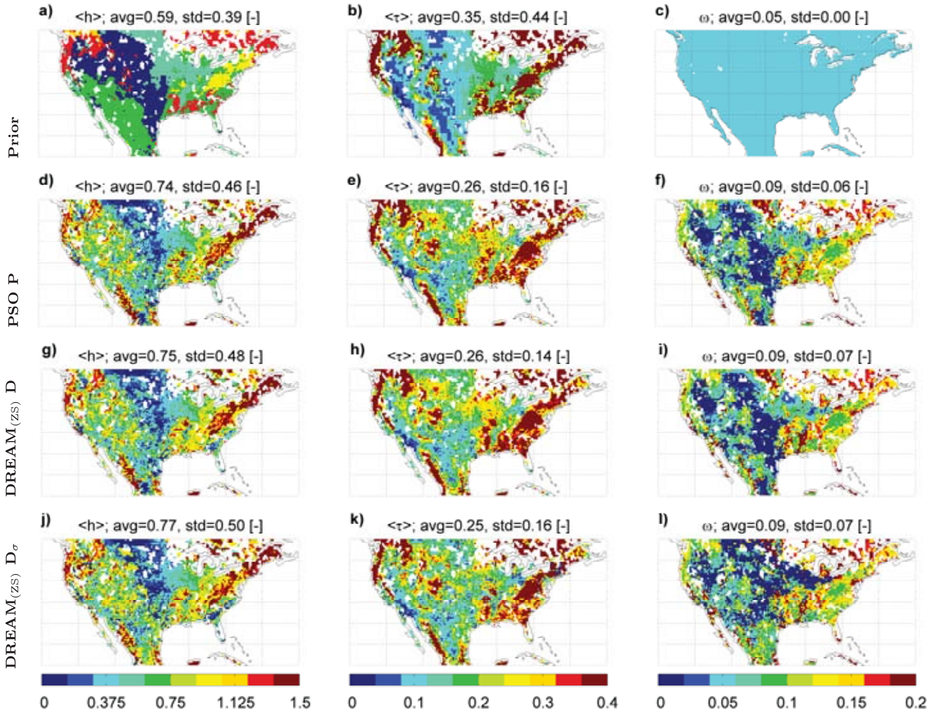

Figure 3: Parameter values for (left) $< h >$, (middle) $< \tau >$, and (right) ω, for the (top row) prior distribution and scenarios (second row) P, (third row) D and (fourth row) D_σ.

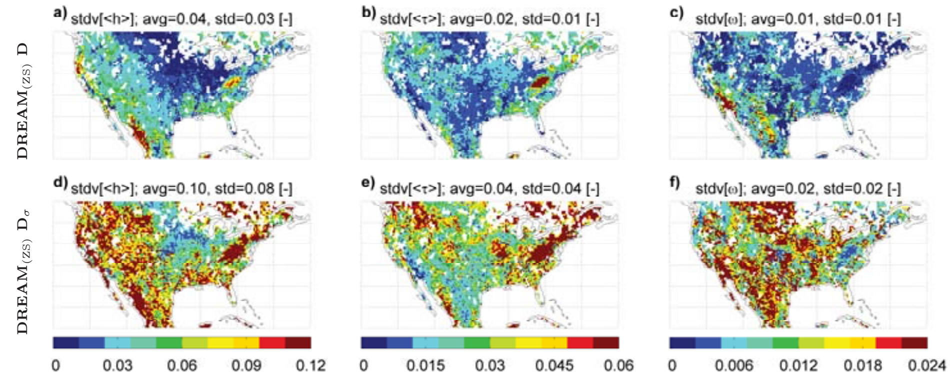

Figure 4: Uncertainty in parameter estimates for (left) $< h >$, (middle) $< \tau >$, and (right) ω, obtained with DREAM$_{\text{(ZS)}}$ scenario (top row) D and (bottom row) D$_\sigma$.

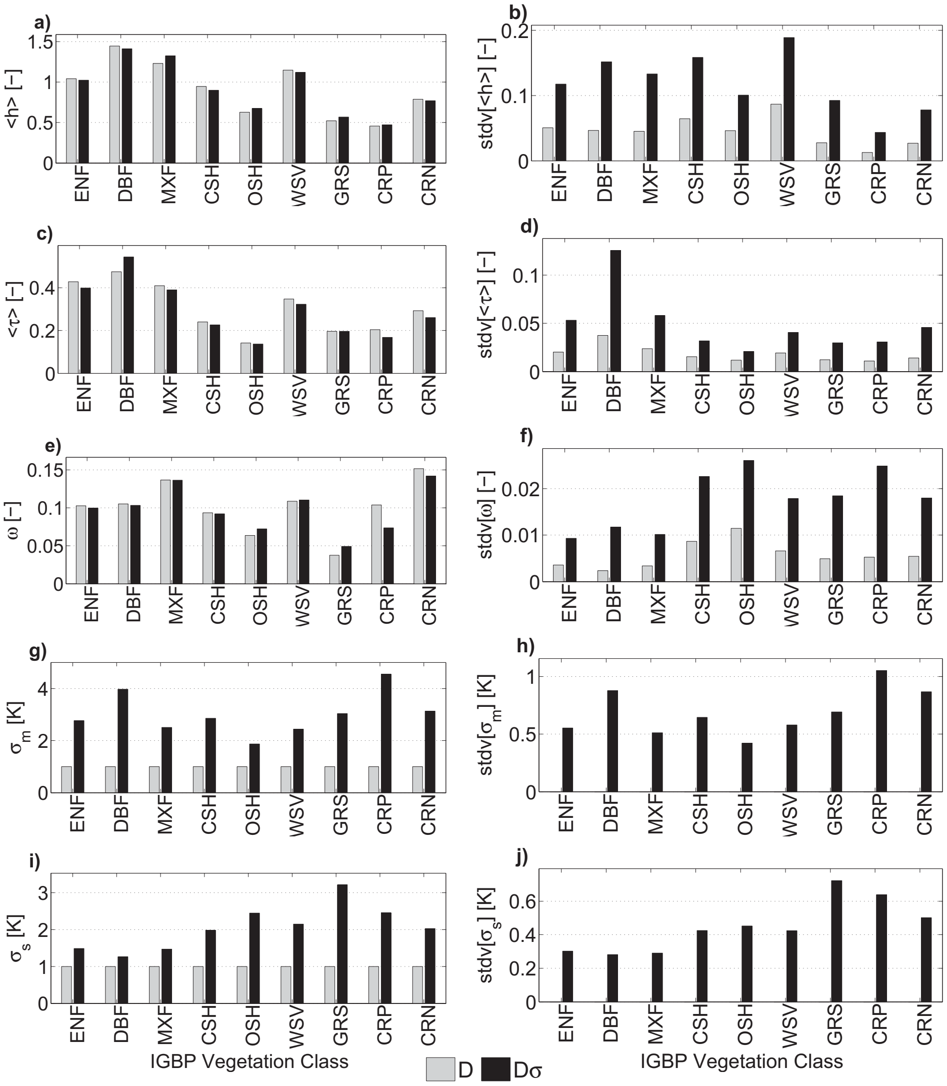

Figure 5: (Left) MAP parameter values and (right) uncertainties aggregated per vegetation class for DREAM$_{(ZS)}$ scenarios D and D$_\sigma$. Each row represents a different parameter: (a,b) $< h >$, (c,d) τ, (e,f) ω, (g,h) σ_m, (i,j) σ_s.

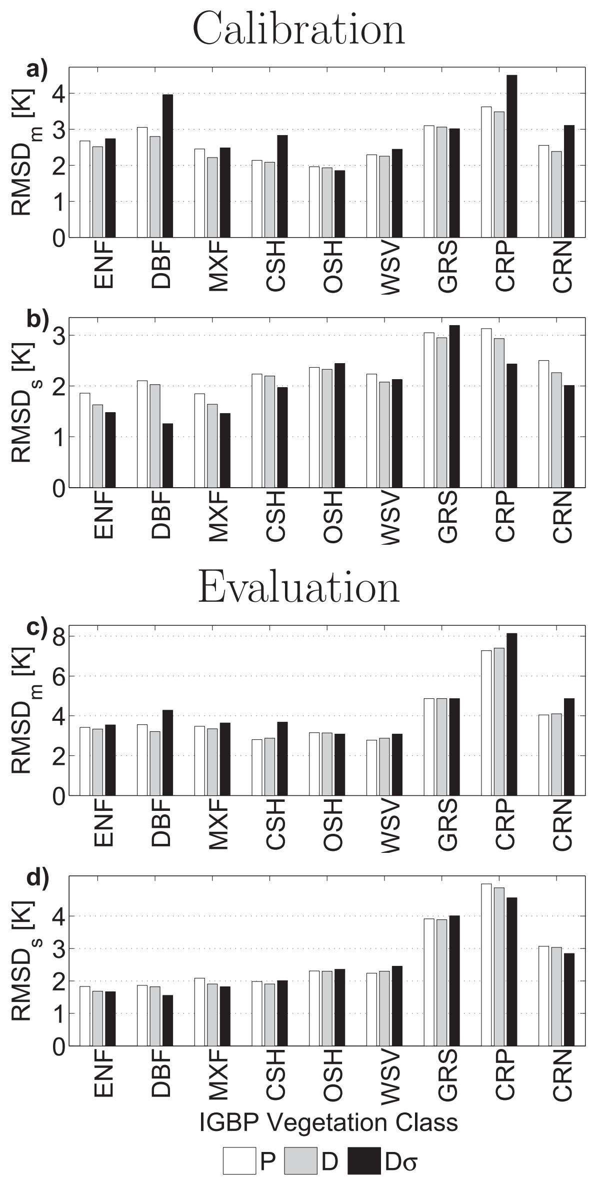

Figure 6: RMSD in long-term Tb (a,c) average and (b,d) standard deviation during the (top) calibration (1 July 2011 - 1 July 2012) and (bottom) evaluation period (1 July 2010 - 1 July 2011), using the MAP parameter values derived from PSO (scenario P) and DREAM$_{(ZS)}$ (scenarios D and D$_\sigma$).

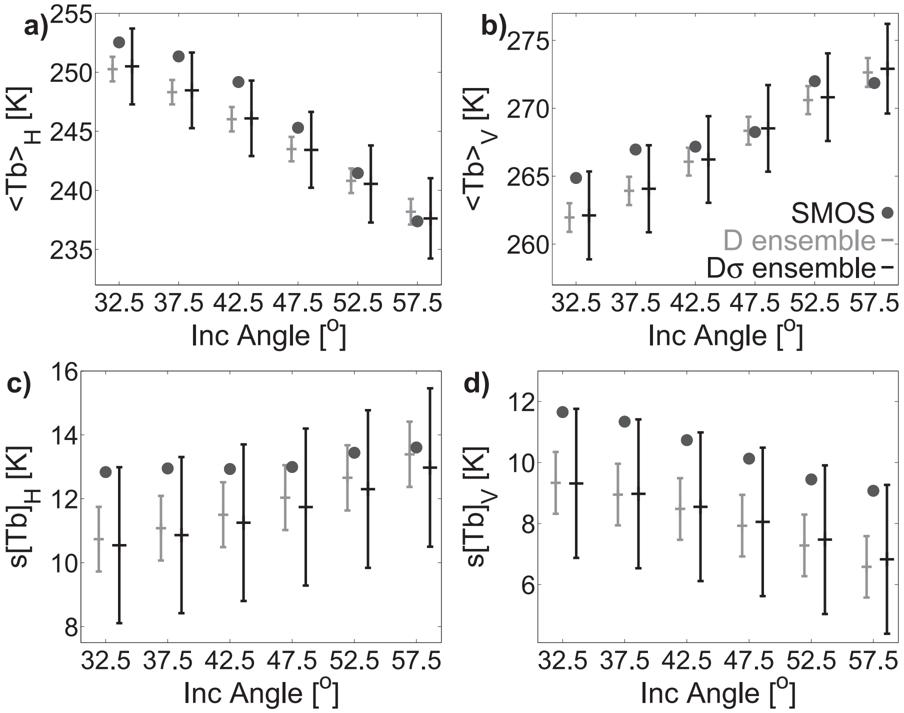

Figure 7: (a-b) Long-term average and (c-d) standard deviation, for (a-c) H- and (b-d) V-polarized Tb (dots) SMOS observations and (lines) ensemble simulations averaged over the study domain, during the calibration period (1 July 2011 - 1 July 2012) and only including ascending time steps. The simulations use an ensemble of parameter estimates derived with DREAM$_{(ZS)}$ scenarios (gray) D and (black) D$_\sigma$. The ensemble mean is shown by a central horizontal dash. The error bars indicate the total simulation and observation uncertainty and are drawn around the simulated Tb for illustration. For clarity, symbols are slightly offset from the nominal incidence angle.

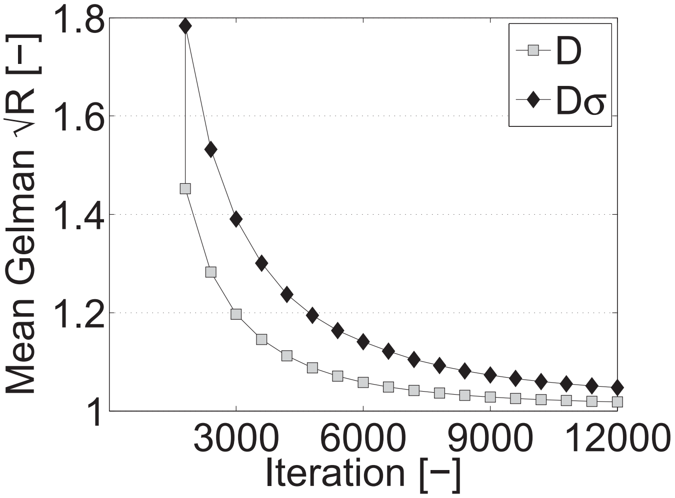

Figure 8: Gelman-Rubin convergence diagnostic $\sqrt{\mathrm{R}}$ for the two DREAM$_{\mathrm{(ZS)}}$ MCMC simulation scenarios. The metric is averaged over all calibrated parameters, and across the study domain.