Spatial horizontal correlation characteristics in the land data assimilation of soil moisture

X. Han1,2, X. Li1, H. J. Hendricks Franssen2, H. Vereecken2, and C. Montzka2
1Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, Gansu, 730000, China
2Forschungszentrum Jülich, Agrosphere (IBG 3), Leo-Brandt-Strasse, 52425 Jülich, Germany

Correspondence to: X. Li (lixin@lzb.ac.cn)

Received: 12 September 2011 – Published in Hydrol. Earth Syst. Sci. Discuss.: 27 September 2011
Revised: 12 April 2012 – Accepted: 27 April 2012 – Published: 10 May 2012

Abstract. Remote sensing images deliver important information about soil moisture, but often cover only part of an area, for example due to the presence of clouds or vegetation. This paper examines the potential of incorporating the spatial horizontal correlation characteristics of surface soil moisture observations in land data assimilation in order to obtain improved estimates of soil moisture at uncovered grid cells (i.e. grid cells without observations). Observing system simulation experiments were carried out to assimilate the synthetic surface soil moisture observations into the Community Land Model for the Babaohe River Basin in northwestern China. The estimation of soil moisture at the uncovered grid cells was improved when information about surrounding observations and their spatial correlation structure was included. Including an increasing number of observations for covered and uncovered grid cells in the assimilation procedure led to a better prediction of soil moisture with an upper limit of five observations. A further increase of the number of observations did not further improve the results for this specific case. High observational coverage resulted in a better assimilation performance, depending also on the spatial distribution of observation data. In summary, the spatial horizontal correlation structure of soil moisture was found to be helpful for improving the surface soil moisture data characterization, especially for uncovered grid cells.

1 Introduction

The estimation of soil-atmosphere exchange fluxes of water and energy can be improved by assimilating in situ measurements and remote sensing products (Li et al., 2007; Moradkhani, 2008; Reichle et al., 2008). Li et al. (2010), Liu et al. (2011), Reichle et al. (2007), Tian et al. (2010) and Yang et al. (2009) proved that the assimilation of real passive microwave remotely sensed surface soil moisture or brightness temperature could improve the model estimation of soil moisture. The soil moisture products of scatterometer were also used in data assimilation (Dharssi et al., 2011; Mahfouf, 2010; Brocca et al., 2010b). Pan and Wood (2010) analyzed the impact of satellite spatial availability on the soil moisture assimilation in a synthetic experiment. Montzka et al. (2011) and Nie et al. (2011) showed that soil moisture observations could be used to optimize the parameters of a land surface model. Ghent et al. (2010) showed that the assimilation of surface temperature from three remote sensing products could improve the modeled surface temperature, soil moisture, latent and sensible heat fluxes. Reichle et al. (2010) also studied the benefit of surface temperature assimilation in improving the model estimation of surface temperature. Xu et al. (2011) showed that the characterization of the sensible and latent heat fluxes was improved by the assimilation of remotely sensed surface temperature.
These studies showed that remotely sensed surface soil moisture and surface temperature have become the important data sources in regional land data assimilation applications. Examples are the surface soil moisture products of the microwave sensors AMSR-E (Advanced Microwave Scanning Radiometer for EOS) (Njoku and Chan, 2006), ASCAT (Advanced Scatterometer) (Naemi et al., 2009), SMOS (Soil Moisture and Ocean Salinity) (Kerr et al., 2010), the upcoming SMAP mission (Soil Moisture Active Passive) (Entekhabi et al., 2010), and the land surface temperature of thermal infrared sensor MODIS (Moderate Resolution Imaging Spectroradiometer) (Wan and Li, 2008). However, soil moisture retrieval based on microwave measurements is often hampered by the presence of vegetation cover (Dorigo et al., 2010; Njoku and Chan, 2006) or topography (Flores et al., 2009; Matzler and Standley, 2000; Pellenc et al., 2003). Moreover, passive microwave sensor records can be contaminated by radio frequency interferences and accordant pixels have to be excluded from the analysis (Anterrieu, 2011; Skou et al., 2010). The land surface temperature retrievals can also be influenced by the cloud cover (Coll et al., 2009; Wan and Li, 2008). All these effects may lead to incomplete spatial coverage of the soil moisture and surface temperature fields from remotely sensed measurements for the study area of interest leading to an exclusion of these areas for data assimilation purposes as no data are available. However, land surface variables like soil moisture or surface temperature show a spatial horizontal correlation (we will use “spatial correlation” as a synonym for “spatial horizontal correlation”). The presence of such correlations implies that land surface variables at the uncovered area can be related to land surface variables at the covered area. Brocca et al. (2010a) and Ryu and Famiglietti (2006) studied the spatial correlation of soil moisture at different spatial scales. The results showed that the spatial correlation pattern of soil moisture could be modeled with the help of geostatistical approaches and the regional soil moisture content could be estimated using a fixed number of samples. This provides the opportunity to improve the estimation of land surface variables by propagating the information of observations from covered areas to uncovered areas. De Lannoy et al. (2009) and Reichle and Koster (2003) have studied the impact of horizontal error correlation in the model forecast covariance matrix for soil moisture assimilation. On the other hand, due to the inaccuracies in the spatial registration of remote sensing products (Townshend et al., 1992), and the difference between the spatial resolution of remote sensing platforms and the land surface model, remotely sensed observations cannot be mapped directly onto the model grid cell. This means it is difficult to find the observation data located at the same spatial location as the model grid cell. There are offsets between the observation location and the model location.

In land data assimilation, these two common problems should be considered jointly: uncovered model grid cells because of the observational spatial coverage and the distance between the observation location and model location because of the inaccuracies in the spatial registration of remote sensing products. We need to assess whether the neighboring surrounding observations can be used to improve model estimations or not. In order to incorporate the spatial correlation pattern of observations into a data assimilation system, we chose the local observation selection technique (Greybush et al., 2010; Houtekamer and Mitchell, 1998; Hunt et al., 2007; Whitaker et al., 2008), in which the observations near a model grid cell can be used for updating with data assimilation.

The local observation selection limits the impact of observations further away from the grid cell by filtering out the small correlations associated with these observations. The benefits and performance of the local observation selection were discussed by Greybush et al. (2010) and Hunt et al. (2007) in the framework of Local Ensemble Transform Kalman Filter (LETKF). Presently, geostatistical methods which can characterize the spatial correlation structure of state variables such as soil moisture have not yet been used in the local observation selection technique of data assimilation. In particular, the geostatistical semivariogram model can be used to model the horizontal spatial dependence among observations (Chiles and Delfiner, 1999; De Lannoy et al., 2006; Lakhankar et al., 2010).

The objective of this study is to evaluate the potential of incorporating the spatial correlation characteristics of observations into LETKF in order to improve the soil moisture profile estimation. The organization of this paper is as follows. Section 2 presents a review of the study area, the Community Land Model, as well as the details of model input data. Section 3 presents the experimental design and the explanation of the methodologies used. Results in Sect. 4 are derived from the observing system simulation experiments. Section 5 provides a brief summary and discussion of the key results.

2 Study area, model and data

2.1 Study area

The study area was the Babaohu River Basin, which is a sub-catchment in the upper reaches of the Heihe River Basin, China’s second largest inland river basin. The elevation is from 3000 m to about 5000 m. The main land cover is grassland. The warm season is from April to September. The annual mean air temperature, annual precipitation and annual potential evapotranspiration are 0.7°C, 391.1 mm and 1529.8 mm, respectively. Permafrost and glaciers are widely present in this area, and snow prevails in winter (Li et al., 2009). Figure 1a shows the digital elevation model of the Babaohu River Basin.
2.2 Model and input data

The new developed Community Land Model 4 (CLM) (Oleson et al., 2010) was used to describe the soil-vegetation-atmosphere transfer of water, energy and matter. CLM represents several aspects of the land surface including surface heterogeneity and consists of components or submodels related to land biogeophysics, hydrologic cycle, biogeochemistry and ecosystem dynamics (Oleson et al., 2010). The model of the Baobaohe River Basin consists of 15 soil layers, and the model grid resolution is 1 km. There are 3640 active grid columns.

The 25 km and 3 hourly atmospheric forcing data from the Global Land Data Assimilation System (GLDAS) project (Rodell et al., 2004), in which several forcing products from the model reanalysis and remote sensing were combined together, were interpolated on a 1 km grid with a temporal resolution of 1 h. Precipitation was interpolated in time using a linear disaggregation method provided by the Global Soil Wetness Project 2 (GSWP2). The temporal interpolations of incident solar radiation, incident longwave radiation, wind speed, relative humidity, air pressure and air temperature were based on the cubic spline method (Dai et al., 2003). The quasi-physically based high resolution meteorological interpolation model MicroMet, in which the spatial interpolations are performed using the Barnes objective analysis scheme (Liston and Elder, 2006) and the 1 km SRTM Digital Elevation Model were used in the spatial interpolation of the forcing data.

We used the MODIS 500 m Plant Functional Type (PFT) scheme of product MCD12Q1 to replace the CLM 0.5° PFT data. The MODIS PFT was projected into Longitude-Latitude Projection and resampled to 1 km using MRT (MODIS Reprojection Tool). Next, the MODIS PFT (12 types) was translated to CLM PFT (17 types) using the 1 km global climate data WorldClim (Hijmans et al., 2005) and the methods proposed by Bonan et al. (2002). The mean temperature for the warmest and coldest season of the year, the annual precipitation, the precipitation of the driest, warmest and coldest seasons are necessary input data for the PFT translation and were taken from the WorldClim database. Additionally, the growing-degree days above 5°C, calculated from the GLDAS forcing data, are necessary input to reclassify MODIS PFT to CLM PFT (Bonan et al., 2002). The 4-day MODIS leaf area index (LAI) data (MCD15A3) at 1 km resolution was used to calculate the monthly LAI, and the stem area index (SAI) was estimated based on the methods proposed by Lawrence and Chase (2007).

In CLM, the soil color determines dry and saturated soil albedo. The soil thermal and hydraulic properties are estimated from sand, clay, and organic matter content. The maximum fractional saturated area is used to determine surface runoff and infiltration. The 1 km soil color data was calculated with help of the method of Lawrence and Chase (2007), and the soil texture data. The soil sand fraction, clay fraction, organic matter density and soil bulk density were from the Harmonized World Soil Database v1.1 (HWSD) (FAO,
3.2 Spatial correlation and geostatistics

The presence of horizontal spatial dependence of land surface properties is typically identified with help of a semivariogram analysis (Goovaerts, 1997; Ryu and Famiglietti, 2006). Several semivariogram models are often used to characterize the spatial dependence such as the Gaussian model, the exponential model, the spherical model and the Matern model (Goovaerts, 1997; Minasny and McBratney, 2005). In this paper, we only considered these semivariogram models to characterize the spatial dependence of surface soil moisture. These models are defined as follows:

\[
\gamma(h) = c_0 + c \left( 1 - \exp \left( -\frac{h}{a} \right) \right) \quad \text{spherical (1)} \\
\gamma(h) = c_0 + c \left( 1 - \exp \left( -\frac{3h^2}{a^2} \right) \right) \quad \text{exponential (2)} \\
\gamma(h) = c_0 + c \left( 1 - \exp \left( -\frac{15h^2}{a^2} \right) \right) \quad \text{Gaussian (3)} \\
\gamma(h) = c_0 + c \left( \frac{1}{2 \nu - 1} \frac{h}{\alpha} \right) \frac{2^{\nu/2}}{\Gamma(\nu)} K_\nu \left( \frac{2^{\nu/2} h}{\alpha} \right) \quad \text{Matern (4)}
\]

where \(c_0\) is nugget, \(c\) is equal to sill minus nugget, \(h\) is the separation distance, \(a\) is the (effective) range parameter, \(K_\nu\) is a modified Bessel function of second order, \(\Gamma\) is the gamma function and \(\nu\) (kappa) is called “smoothness parameter” \((\nu > 0)\). The normalized semivariogram \(\gamma(h)_{\text{Nor}}\) is defined as \(\gamma(h)_{\text{Nor}} = \gamma(h)/(c_0 + c)\) and the correlogram is then given by \(1 - \gamma(h)_{\text{Nor}}\). The nugget value describes the unresolved variance at the scale smaller than the smallest lag distance whereas the sill describes the variance of the observed data as they become spatially independent. The calculated correlogram is normalized using the maximum correlogram value, which means that the correlogram value is equal to one at the observation locations. It gradually reduces towards 0.0 as the distance from the analysis grid cell increases, and is null beyond the specified correlation range. The observation selection scheme in LETKF is as follows: spatial correlations between the model location and the observation locations are calculated according to the fitted semivariogram; the observations whose correlograms are larger than a predefined threshold are selected for each model grid cell.

In order to represent the vegetation or cloud cover impacts on the observation data, we derived the mask data using the valid grid cells from the MODIS land surface temperature products of Terra (MOD11A1- Daytime) on 13 April 2008, 1 May 2008 and 26 June 2008, respectively. In the assimilation experiments, it is assumed that only for the valid grid cells from the remote sensing image is meaningful information available, whereas for the other grid cells due to cloud cover and vegetation no meaningful information could be extracted. The positions of the extracted valid grid cells are used as the mask and the synthetic observation data located at these grid cells were selected as the observation data to be assimilated. Assimilation experiments with three different masks were performed which are shown in Fig. 1b–d, respectively. In total 2559, 1697 and 1143 valid observation grid
Fig. 2. Flow chart of data assimilation procedure for covered and uncovered grid cells. “Obs_Num” is the number of observations.

3.3 Local ensemble transform Kalman filter

The local ensemble transform Kalman filter proposed by Hunt et al. (2007) is a new variant of the ensemble Kalman filter (EnKF). The main difference compared to the various kinds of implementations of EnKF, such as the stochastic EnKF (Burgers et al., 1998), the deterministic methods (Whitaker and Hamill, 2002), the ensemble adjustment Kalman filter (EAKF) (Anderson, 2001), the ensemble transform Kalman filter (ETKF) (Bishop et al., 2001) and the Ensemble Square Root Filter (EnSRF) by Whitaker and Hamill (2002), is the local analysis scheme used by LETKF (Houtekamer and Mitchell, 1998; Hunt et al., 2007; Kalnay et al., 2007; Miyoshi and Yamane, 2007; Whitaker et al., 2008).

Here, the spatial correlation characteristics of the observations were considered in the framework of the LETKF local analysis scheme. It only considers the observations located in a local region surrounding the analysis model grid cell. The observations were selected for each grid cell before data assimilation. The grid cell by grid cell analysis method can easily be parallelized and is useful to decrease the computational burden for the large scale data assimilation. For the detailed introduction of LETKF, please refer to Hunt et al. (2007).

3.4 Ensemble generation

The uncertainties in the model calculations were represented by the perturbed forcing data, soil parameters and the initial data (from the spin-up period). The precipitation data were multiplied by log-normal distributed noise (mean = 0, std = 0.7) (Yilmaz et al., 2011). We generated the random noises of shortwave radiation, longwave radiation and air temperature using a multi-normal distribution, in which the cross correlation errors were imposed with mean of ([0, 0, 1.0, 0.0]) and standard deviation matrix of ([2.0, 0.4, 0.4], [0.4, 0.3, -0.6], [0.4, -0.6, 20]). This cross correlation was based on the atmospheric balance between radiation, clouds, and air temperature (Reichle et al., 2010). Using this multivariate sampling, we can generate physically consistent perturbations like a positive perturbation of the shortwave radiation, a negative perturbation of the longwave radiation and a positive perturbation of air temperature. The random noises imposed on air temperature, shortwave radiation and longwave radiation were additive, multiplicative and additive, respectively. Initial soil moisture was perturbed using an additive normal distributed noise (mean = 0.0, std = 0.02 m$^3$ m$^{-3}$). The soil sand fraction, soil clay fraction, soil organic fraction and maximum fractional saturated area were perturbed using the multiplicative truncated normal distributed noise (mean = 1.0, std = 1.0, lower = 0.8, upper = 1.2). As the perturbation of the soil parameters resulted in an adequate ensemble spread, the model prognostic variables were not perturbed, as was for example done in Kumar et al. (2009). This resulted in an ensemble of model outputs which is the input for the LETKF algorithm.
3.5 Data assimilation strategies

Three assimilation strategies were evaluated, in which different local observation selection options were used to include the spatially correlated observations into the analysis scheme: (1) only one observation was used for each grid cell (1_Obs); only the closest observation was included; (2) no more than five observations were used for each grid cell (5_Obs); in addition to the closest observation the next closest four observations were included; (3) no more than nine observations were used for each grid cell (9_Obs), eight additional observations were included in the data assimilation procedure. Figure 2 summaries the data assimilation procedure for the covered and uncovered grid cells.

The above three strategies were used for both covered grid cells and uncovered grid cells. All uncovered grid cells with sufficient correlated observations in the neighborhood were updated, and the model grid cells without sufficiently correlated observations were not updated. The impacts of three different observational coverages (as represented by the three different masks) on the assimilation results were also analyzed.

The soil moisture assimilation was only concerned with updating the liquid soil moisture content and the synthetic observations were also generated from the liquid soil moisture. The update of soil moisture content was constrained such that the sum of liquid soil moisture content and ice content was never larger than the soil porosity. The soil moisture of the surface soil layer (5 cm depth) was updated using the corresponding observations. The soil moisture contents of the other nine unobserved lower soil layers were updated based on the cross correlation between the observed upper soil layer and the unobserved layer.

It should be noted that some of the observations used in the analysis at one grid cell are also used in the analysis of another neighboring grid cell. This imposes a smoothing effect from one grid cell to the next. The threshold value of correlogram to include observations was set equal to 0.1; it means that all the observations with a correlogram larger than 0.1 were considered to be included in the data assimilation. The small threshold of 0.1 guarantees that almost all grid cells have their associated correlated observations. In this study, the spatial correlogram values of the nine observations range...
Fig. 4. Mean RMSE values of soil moisture (including 95% confidence intervals) for open loop simulations (CLM) and three assimilation strategies of 1 observation used (1 Obs), 5 observations used (5 Obs) and 9 observations used (9 Obs) for 50 cm and 80 cm depth. Results are shown for 70% – 50 cm (a), 80 cm (b) – 47% – 50 cm (c), 80 cm (d) – and 31% – 50 cm (e), 80 cm (f) – observational coverages.

from 0.5 to 1.0. Of course in case of only weak correlations, the correcting influence of observations is limited and those weakly correlated observations hardly will reduce the variance.

4 Results and discussion

In order to evaluate the quality of the obtained results, the Root Mean Square Error (RMSE) and the Nash-Sutcliffe model efficiency (NSE) coefficient were calculated. Because RMSE values are usually affected by the mean bias or the mean variation differences, we added the NSE as another measure (Reichle et al., 2010). NSE values represent the correlation between the estimation and the observation. RMSE and NSE values for each grid cell were calculated according to:

\[ \text{RMSE} = \sqrt{\frac{\sum_{n=1}^{N} (\text{Estimated} - \text{Truth})^2}{N}} \]  

\[ \text{NSE} = 1 - \frac{\sum_{n=1}^{N} (\text{Estimated} - \text{Truth})}{\sum_{n=1}^{N} [\text{Truth} - \text{mean(Truth)}]^2} \]  

where “Estimated” is the ensemble mean of open loop run or the ensemble mean after assimilation. \( N \) is the number of time steps, it is 2208 for this study. The smaller the RMSE value and the larger the NSE value are, the better the assimilation results will be.

First we calculated the RMSE and NSE values of each grid cell in the whole time window. Then the mean RMSE and mean NSE for the covered and uncovered grid cells were calculated separately. The 95% confidence intervals of the mean RMSE (NSE) were also calculated on the basis of the RMSE (NSE) values at individual grid cells using the Bayesian methods (Oliphant, 2006).

Figure 3 shows the mean soil moisture RMSE values for open loop simulations and the assimilation results under different amounts of local observations (1, 5 or 9), both for covered and uncovered grid cells at 10 cm depth (Fig. 3a, c and e).
Fig. 5. Mean NSE values of soil moisture (including 95% confidence intervals) for open loop simulations (CLM) and three assimilation strategies of 1 observation used (1_Obs), 5 observations used (5_Obs) and 9 observations used (9_Obs) for 10 cm and 30 cm depth. Results are shown for 70% – 10 cm (a), 30 cm (b) – 47% – 10 cm (c), 30 cm (d) – and 31% – 10 cm (e), 30 cm (f) – observational coverages.

The basin scale mean RMSE values and mean NSE values for all grid cells (including all covered and uncovered grid cells) can give us additional insight in the performance of the different assimilation strategies. Figure 7 shows the mean basin scale soil moisture RMSE values for open loop simulations and the other three assimilation results of 1, 5 and 9 local observations used at all grid cells for 10 cm (Fig. 7a), 30 cm (Fig. 7b), 50 cm (Fig. 7c) and 80 cm depth (Fig. 7d). These results are presented for 70%, 47% and 31% observational coverages. Figure 8 shows the corresponding mean basin scale soil moisture NSE values.

The RMSE values reported in Fig. 7 illustrate again that 5_Obs is the best assimilation strategy. A comparison of the results for the 5_Obs strategy for different observational coverages illustrates that the results for 70% observational coverage are superior to the other ones. Surprisingly, the results for 31% observational coverage are slightly better than the ones for 47% observational coverage. The results for the NSE values (Fig. 8) are consistent with the ones for Fig. 7.

In order to compare the spatial distribution of changes of RMSE values, we show the basin scale 10 cm depth soil
Fig. 6. Mean NSE values of soil moisture (including 95% confidence intervals) for open loop simulations (CLM) and three assimilation strategies of 1 observation used (1_Obs), 5 observations used (5_Obs) and 9 observations used (9_Obs) for 50 cm and 80 cm depth. Results are shown for 70% – 50 cm (a), 80 cm (b) – 47% – 50 cm (c), 80 cm (d) – and 31% – 50 cm (e), 80 cm (f) – observational coverages.

The results for the surface soil moisture characterization illustrate that the estimations for covered and uncovered grid cells can be improved when for a certain grid cell surrounding observations were included in the estimation. The estimations at the uncovered grid cells were also improved when the surrounding correlated observations were used. Results are better if more observations were included in the data assimilation, but more than five observations of surface soil moisture did not much improve the estimations in this case.

The lower soil layers (10 cm, 30 cm, 50 cm and 80 cm) were also updated based on the cross correlation between the surface soil layer and lower soil layer in the LETKF. The improved results in the lower layers prove that the vertical correlations among the soil layers can be useful to transfer the surface observation to the lower layers. The improved results at the uncovered grid cells also prove that the horizontal correlations among the observation data can be useful to transfer the observation data from the observed grid cells to the unobserved grid cells. Therefore, the combination of the vertical correlations and the horizontal correlations in the soil moisture data assimilation is helpful for improving the characterization of soil moisture heterogeneity.

In addition to including the spatial dependence of observed state variables, it may be worthwhile to also include temporal dependence of land surface observations in data assimilation schemes for an improved estimation of state variables. Experimental data clearly shows that land surface variables, such as soil moisture and soil temperature are correlated...
Temporal stability of soil moisture (Brocca et al., 2010a; De Lannoy et al., 2006) has been extensively studied. Dunne and Entekhabi (2005) studied the ensemble Kalman smoother in the soil moisture assimilation, but the potential of temporal correlation has probably not yet been fully explored in the framework of data assimilation schemes. Moreover, land cover and soil type-dependent spatial correlations may improve the findings of this study.

The spatial interpolation methods for the forcing data and the spatial resolution of the input data also influence the spatial correlation characteristics of the model. The 1 km spatial resolution we chose in this study is finer than the common...
This figure shows the basin scale 10 cm soil moisture RMSE values for open loop simulations (a) and three assimilation strategies of 1 observation used (1_Obs) (b), 5 observations used (5_Obs) (c) and 9 observations used (9_Obs) (d) in case of 70% observational coverage.

spatial correlation used in the land data assimilation. The results of this work rely on the specific spatial correlation characteristics of input data. The spatial correlation range of high resolution data is shorter than that of coarse spatial resolution data. It is necessary to study what could be the contribution of spatial correlation for soil moisture assimilation in (very) high resolution land surface models. The synthetic experiment typically overestimates the performance of the method because conceptual uncertainty and model uncertainty are neglected from the analysis (Kumar et al., 2009). Therefore, a real-world case study with more data is desirable.

The masks used in this study cannot represent the true spatial coverage of microwave sensors, which is easily influenced by the spatial distribution of vegetation. More than 90% of the grid cells in this study area are covered by short grass and we think that the masks used in this work are more suitable for representing the proposed idea to evaluate the impacts of spatial correlations.

An operational application of the presented spatial correlation patterns may be feasible for the assimilation of surface soil moisture in weather forecast models like the ECMWF Integrated Forecast System (Drusch et al., 2009). However, the computational issue for the large scale high resolution data assimilation will be challenging, because we have to fit the semivariograms for soil moisture at each time step. The fitting algorithm will increase the computation remarkably.

5 Summary

We carried out a set of synthetic observing system simulation experiments to explicitly include the spatial correlation characteristics of soil moisture in the land data assimilation for updating surface soil moisture at locations without remote sensing observations. Two different spatial correlation schemes were evaluated: (1) for the grid cells covered with observations, additional observations around the grid cells were included in the data assimilation; (2) for the grid cells without observations, observations surrounding these grid cells were used. First, we characterized the spatial correlations of soil moisture on the basis of a geostatistical semivariogram analysis. Second, the fitted semivariogram models and the corresponding correlogram values were used to choose the correlated observations for each model grid cell in LETKF during assimilation. Third, the selected observations were sorted according to their correlogram values and the simulation experiments considered different numbers of observations to be used in the data assimilation.
Our study showed that including the horizontal spatial dependence of surface soil moisture observations in the data assimilation scheme improved the estimation. The model estimations at the grid cells without observations can be improved using the geostatistical model and LETKF. In this particular case, best results were obtained by assimilating five observations. The best assimilation performance was achieved for a high observational coverage, but the results also depend on the spatial distribution of the observation data. For example, the results for 47% observational coverage are worse than the results for 31% observational coverage, because the spatial distribution for the 31% observational coverage is more uniform than the one for the 47% observational coverage: for the 47% observational coverage we can find a large gap between the patches of observations (Fig. 1c). Therefore, the lower observational coverage can be offset by the horizontal spatial correlation.

In summary, the proposed spatial correlation method will be helpful to the soil moisture data assimilation in two ways: increasing the spatial availability of observations and offsetting the impact of spatial registration error in the remote sensing data processing. Since the spatial correlations are the intrinsic characteristics of many land surface variables, it is worth extending the proposed methods to the land data assimilation of other observation data.

Acknowledgements. This work was supported by the National High-tech Program (863) of China (grant number: 2009AA122104), the Knowledge Innovation Program of the Chinese Academy of Sciences (grant number: KZCX2-EW-312) and the NSFC (National Science Foundation of China) project (grant number: 40901160, 40925004). Carsten Montzka received financial support from the DFG (German Science foundation) by means of the project TR-32 “Patterns in Soil-Vegetation-Atmosphere systems: Monitoring, modeling and data assimilation”, which is gratefully acknowledged. The data used in this study were acquired as part of the mission of NASA’s Earth Science Division and archived and distributed by the Goddard Earth Sciences (GES) Data and Information Services Center (DISC).

Edited by: Y. Li

References


X. Han et al.: Spatial horizontal correlation characteristics in the land data assimilation of soil moisture


