Hydrol. Earth Syst. Sci. Discuss., 8, C5327–C5332, 2011
www.hydrol-earth-syst-sci-discuss.net/8/C5327/2011/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Spatial horizontal correlation characteristics in the land data assimilation of soil moisture and surface temperature" *by* X. Han et al.

X. Han et al.

hanxj@lzb.ac.cn

Received and published: 17 December 2011

Response to reviewer #1 comments on "Spatial Horizontal Correlation Characteristics in the Land Data Assimilation of Soil Moisture and Surface Temperature"

Thanks for your comments and recommendations to help us improving the presentation of this study and the organization of the manuscript. Please find below our responses (in blue):

Major Comments:

Comment: In this study, ground truth and synthetic observations are simulated using CLM model at a resolution of 1 km with interpolated forcing data and land surface data, such as soil type. Both soil moisture and surface temperature are heavily related to atmospheric inputs and soil types. The hourly 1 km atmospheric forcing data are derived by interpolating GLDAS data, which are at 25 km and 3 h resolutions. The spatial patterns of the 1 km forcing data are significantly influenced by the data interpolating method, which means the spatial correlation characteristics of simulated soil temperature and surface temperature are unavoidably affected by the data interpolating method. In addition, the upper reach of the Heihe Basin is a data-sparse area. There may be a very limited number of soil samples over this area. So, the HWSD soil data may not be able to represent the variability of soil types at the resolution of 1 km over this area. Due to these two reasons, I would argue how much of the spatial correlations of your data (truth and synthetic observations) account for the characteristics of real soil moisture and surface temperature. Please justify the rationale of your data selection.

Response: The upper reach of the Heihe River Basin is a data sparse area. As for the atmospheric forcing data, we use the reanalysis data (GLDAS) to generate the forcing data to be used for the regional modeling. The forcing data can be interpolated using a dynamic model (such as the WRF model) or statistical methods (such as the MicroMet model or bilinear). The HWSD soil data is the only high resolution soil data that we can find for CLM, and it provides us both the soil texture and organic data needed by CLM. The soil data of China in HWSD used the Chinese soil data of the Institute of Soil Science, Chinese Academy of Sciences.

The soil texture (1km), soil color (1km), maximum fractional saturated area (1km), leaf area index (1km) and Plant Functional Type (1km) influence soil moisture and surface temperature. We agree that the spatial interpolation methods influence the spatial correlation characteristics in the model. However, it is important to re-stress that we used a synthetic method to evaluate the proposed methods, in which the truth, the observations and the analysis were based on the same spatial correlation characteristics. Maybe this spatial correlation differs from the true spatial correlation pattern detected in this area, but it is the real correlation contained in this synthetic experiment for this specific synthetic observation. In addition, we agree that the spatial correlation will be different at different spatial scales and for different areas. With help of data assimilation, this spatial correlation can be estimated dynamically with the observations used in the assimilation. We argue that the estimated spatial correlation characteristics from the specific observations will be the representation of the synthetic correlation of this study case.

The reviewer addresses the important question to what extend the results for this synthetic study can be extrapolated then to real-world conditions. This question will be discussed in our new discussion section. We will argue that the forcings and soil data used in this synthetic study are not expected to show a (much) stronger spatial autocorrelation than we expect in reality. However, we also agree with the reviewer that this would have to be investigated further for a real-world case with more data.

Comment: In your experiment design, simulated soil moisture and surface temperature at 09/11/2008 6:00 are selected as the ground truth; simulated soil moisture and surface temperature at 09/11/2007 6:00 are used as model values to be updated. According to the NSE values in table II and table III, the ground truth and the model values are very poorly correlated for both soil moisture and surface temperature. I don't quite understand the rationale for such an experiment design. Could you please provide your reasons in the paper?

Response: We agree that this experiment design was not optimal. So we have redesigned the experiment and used the data from the same period in the data assimilation and extended the assimilation period to three months. We will update the results in the revision.

Comment: In section 3.3, you give a very detailed description about the local ensem-

ble transform Kalman filter. Could you please just briefly introduce the algorithm since it is not the main focus of this paper? You may just describe the part.

Response: Thanks for this suggestion. We will reduce the introduction of LETKF in the revision.

Comment: Could you add more discussions on the threshold of neighboring observations used in data assimilation? According section 4, there are different numbers of neighboring observations for optimal data assimilation of soil temperature and surface temperature. Since the number of neighboring observations is directly related to the spatial correlation characteristic of observation data, it will make your paper more valuable if you dig a little bit deeper.

Response: In the new experimental set-up, we will use one observation (1-Obs), five observations (5-Obs) and nine observations (9-Obs) instead in the data assimilation:

(1) When we only use one observation (1-Obs), only the closest observation is included.

(2) When we use five observations (5-Obs), in addition the four closest observations (with the same distance to the grid cell for which an estimate is required) are included.

(3) When we use nine observations (9-Obs), four further observations are included in the data assimilation procedure.

We will add this discussion in the revision.

Comment: You use spatial correlation as a criterion of selecting neighboring observations for grid with missing values. However, high correlation does not necessarily mean close magnitudes between grids. You may discuss a little bit about this in your paper.

Response: Thanks for this suggestion. We will add this discussion in the revision.

Minor Comments:

Comment: Please pay more attention to the use of the article "the" in your writing.

Response: Thanks. We will improve the English writing in the revision.

Comment: In section 3.1, could you please explain the reason for selecting 10000 meters as the range for generating synthetic observations?

Response: Through the correlation analysis of the soil moisture and surface temperature, we find that the correlation lengths range from several kilometers to hundred kilometers. In this synthetic study, we want to add the spatial correlated noise to the synthetic truth to obtain the corrupted observation. In order to get this spatial correlated noise data, we imposed 10000 meters to represent the synthetic correlated observation noise.

Comment: Please revise lines 251to 253 to make them understandable.

Response: We want to represent the spatial availability of the remote sensing data using the mask. Because of the cloud cover or vegetation effects, we cannot get the fully covered remote sensing data for the study region. So we used one particular MODIS image and obtained its product quality identifier (good or bad) for each grid cell. Then we assumed that for the grid cells with a good product quality a synthetic observation could be used for assimilation, whereas for the other grid cells no information was available.

Comment: From line 371 to line 373, you select 0.001 as the threshold of correlation. I believe that 0.001 is too small for a meaningful correlation. Please justify your selection of this threshold of correlation.

Response: We agree that this threshold is too small. In the revision we have used the value of 0.1. During the data assimilation, we try to find the correlated observations for each grid cell. This small threshold can make sure that almost all grid cells have their associated correlated observations. 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.

Comment: NSE has an upper boundary (1.0). In your discussion, you use "NSE values are xx times larger than" (for example LINE 396). Is it proper to compare NSE values in this way? What is the physical meaning if the NSE value of data A is 2 times larger than that of data B?

Response: We will rewrite the text parts involved and reformulate the sentences where we compare NSE values. We will use the value of NSE and its confidence interval instead.

Comment: Is "spatial horizontal correlation" a professional term? Does it have the same meaning with "spatial correlation"? If yes, you may just use "spatial correlation".

Response: In data assimilation, we have two ways to spread the observation spatially: one is horizontally, the other is vertically. We use the horizontal correlation to transfer the observations to the neighboring grid cells or use the vertical correlation to transfer the observations to deep soil layers at the same grid cell. That is the reason why we used the term "spatial horizontal correlation". However, in the new version of the manuscript we will introduce "spatial horizontal correlation" and clarify that later we will use the term "spatial correlation" as a synonym for spatial horizontal correlation", for sake of brevity.