Reviewer: W. Dorigo (Referee) wouter.dorigo@tuwien.ac.at

We thank Prof. Dorigo for his constructive comments and suggestions (plain text), which significantly improved our paper. Our response is bold text.

The manuscript evaluates a satellite-based ESA CCI soil moisture product and a soil moisture product based on CLM 4.5 simulations using ground observations from approximately 300 meteorological stations in China. Classical metrics are used to evaluate the two products, e.g. Pearson's correlation, bias and the RMSD. In-situ observations are a valuable resource for assessing the performance of satellite and model products but unfortunately the manuscript fails to dig deeper into the causes of observed mismatches. Some of the metrics presented do not make much sense with respect to the evaluation of the ESA CCI SM products improve our dataset understanding at all, e.g. bias and the RMSSD. Especially given the fact that various similar CLM and ESA CCI SM validation papers over China have been recently published, more effort needs to be invested to make this manuscript unique.

My concerns are detailed below.

Major:

1) Several studies comparing soil moisture from land surface models and ESA CCI SM against in-situ observations in China have been recently published (e.g. Lai et al. (2015); Dorigo et al., 2015). To distinguish your work from these studies, you need to put more effort in explaining the differences will real evidence, such as ancillary datasets. In your study, the causes for mismatches observed remain rather speculative.

Response: Based on the comments, we revised the description of the differences between our research and previous studies to make it clearer, please see Page 4 Lines 85−94 and Page 5 Lines 103−114. Previous studies (Albergel et al., 2013; Dorigo et al., 2015) have used only 34 sites for the period 1981–2000 across China and only 20 sites for the period 2008–2010 from the Maqu network in northwest China to investigate the performance of the ESA CCI SM; Lai et al. (2015) validated the temporal variation of soil moisture simulated from the CLM4.0, a previous version of CLM4.5, but used only 30 sites to cover China for the period 1981–1999; they also compared the spatial variation of soil moisture in China using the ESA CCI SM product. However, the performance of the ESA CCI SM product was not discussed in their work. In this study, we conducted an in-depth evaluation of the ESA CCI SM product and CLM4.5 simulation in China using ground-based observations from 306 sites. We investigated their performances over various sub-regions under different climate conditions. This in-depth evaluation provided a better understanding of the quality of both soil moisture products and their potential problems than was hereafter available, and can be used to improve their accuracy.

Moreover, to distinguish our work from these studies, some new discussions about the causes for mismatches were added to Section 4 as suggested, including (1) a detailed description of in situ soil sampling method and a discussion of its potential effect on the

comparison results, (2) the potential effect of in situ precipitation measurements on the comparison results for the CLM4.5 simulation; (3) the impact of the mismatch in soil depths on the statistical metrics. Please see Pages 18−21 Lines 428−520. In addition, in response to the comment, we also revised the analysis of the performances of the ESA CCI SM product and CLM4.5 simulation by using newly defined statistical metrics, including the ubRMSD (with a linear scaling method prior to computing it) and Spearman rank correlation coefficient. Please see Section 2.4 and Section 3.

2) As described at http://www.esa-soilmoisture-cci.org/node/136 and in several publicaions (e.g. Liu et al., 2011, 2012; Dorigo et al, 2015) the methodology to generate ESA CCI SM involves a scaling against GLDAS-Noah to combine scatterometer and radiometer soil moisture products to produce a merged dataset in volumetric units. As a consequence, the mean and dynamic range of ESA CCI SM time series represent those of the GLDAS-Noah surface soil moisture product. Thus, metrics like bias and RMSD mainly reflect differences between GLDAS-Noah and in-situ observations and do not provide much interesting information about the satellite product itself. Therefore, I recommend to exclude them from the analysis and remove all related graphics.

Response: Based on this comment, the statistical metrics used in the old manuscript, including the mean bias (BIAS), root mean square difference (RMSD), normalized standard deviation (SDV), and centered normalized RMSD (E), have been removed; and we mainly focused on the analysis about correlation coefficients and unbiased RMSD (ubRMSD) in the revised manuscript. Please see Section 2.4 (Pages 9−10, Lines 219−237). In addition, the Pearson correlation coefficient was instead by the Spearman rank correlation coefficient (Dorigo et al., 2015), please see Page 10 Lines 235−237; both the ESA CCI SM and CLM4.5 soil moisture datasets were scaled into the dynamic range of the in situ observations using a linear rescaling method (Brocca et al., 2013; Dorigo et al., 2015) prior to computing the ubRSMD (new Eq. 2), please see Page 10 Lines 230−235. Old Fig. 8 (Taylor diagram) and Table 3 have been removed; other old figures and tables related to these metrics were also revised. Please see new Figs. 3−11 and Table 2.

3) It is stated (p5152.line10, p.5163.l15-17) that the high biases for CLM4.5 are caused by inaccurate descriptions of soil characteristics but nowhere in the text evidence is provided. This needs to be elaborated by additional analyses.

Response: Based on the second comment and another reviewer's suggestions, the analysis about the mean bias error (BIAS) and root mean square difference (RSMD) for CLM4.5 has been removed in the revised manuscript. Moreover, because the CLM4.5 simulations were scaled into the dynamic range of the in situ observations using a linear rescaling method based on the mean and standard deviation (Brocca et al., 2013; Dorigo et al., 2015) prior to computing the ubRMSD values, the higher biases over eastern China were not found through the comparison of ubRMSD values. Thus, the related results have been removed in the revised manuscript. However, in response to this comment, we analyzed the effect of soil characteristics on the CLM4.5 simulations by using a new surface dataset

from Shangguan et al. (2012), which was derived by using available observation-based soil characteristics dataset for China. The following figure shows the statistical metrics, including the BIAS and RMSD, for original and new surface datasets over eastern China (China I−IV, defined in Fig. 1 of the manuscript). Results show that new surface dataset (new_surf, blue) reduce biases for CLM4.5 over Eastern China (China I−IV) compared to original one (org_surf, red).

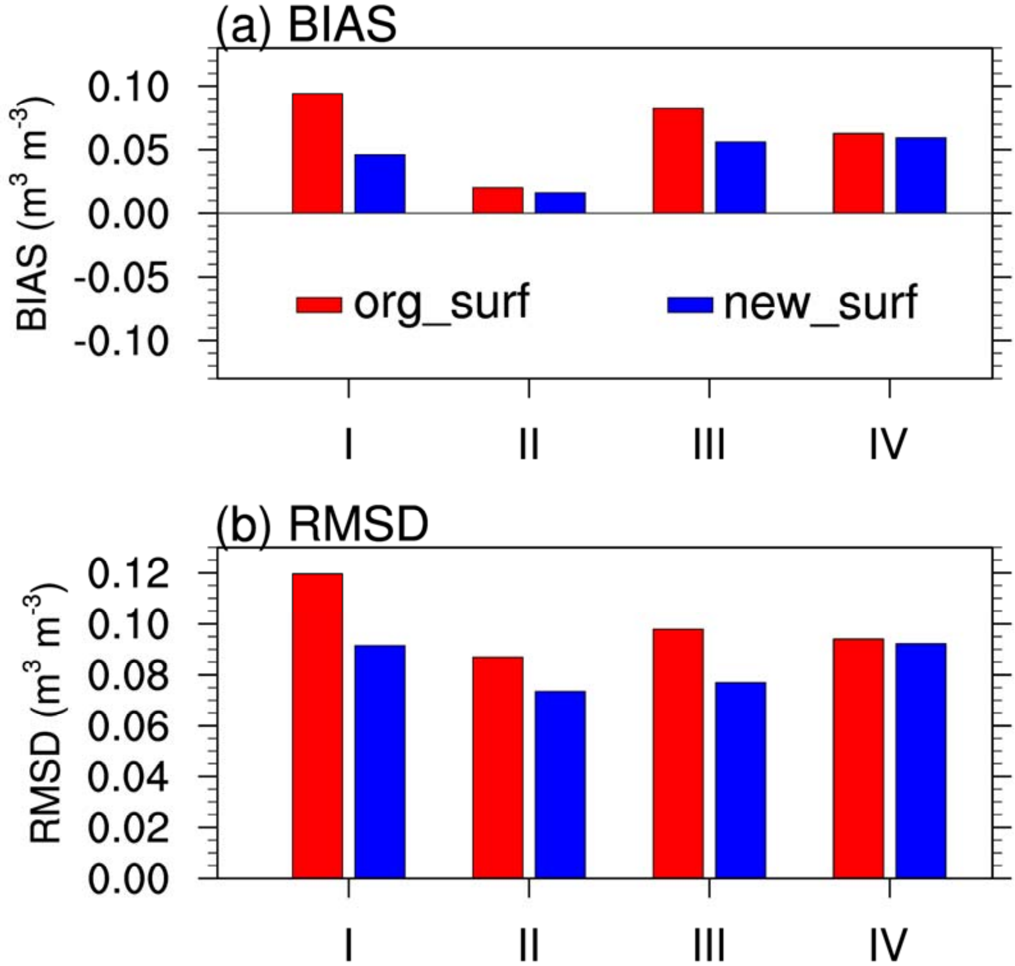

Fig. S1 Average statistical metrics for CLM4.5 using different surface datasets over the four sub-regions of eastern China (see Table 1 of the manuscript): (a) BIAS (CLM4.5 minus in situ observations), (b) RMSD. The original surface dataset (org_surf, red) represents the default ones in CLM4.5 (Oleson et al., 2013) while the new one (new_surf) is from Shangguan et al. (2012).

4) No details are given on how you match the in-situ observations and the gridded products. The ESA CCI SM datasets represents the upper 2 centimetres of the soil, just like the upper layer of the CLM product while, according to your description, the in-situ observations closest to the surface are taken at 10 cm. Thus there is a clear discrepancy in depth which, in particular close to the surface may lead to significantly different dynamics. Therefore, you 1) need to specify which layers and observation depths were compared and 2) what

impact the mismatch between depths may have on your statistics. If there is more than one in-situ station within a grid cell, do you regard the results as individual results? For more details on potential matching strategies see Dorigo et al., 2015.

Response: 1. Based on the comment, we added a detailed description of the match method to Section 2.4: " ***Since soil layer thickness of CLM4.5 model did not match that of in situ observations (0−10 cm), the weighted average was computed based on top four soil layer thicknesses (1.75, 2.76, 4.55, 0.94 cm, respectively)*** **". Please see Page 9 Lines 215−218. How the in situ observations match the gridded products was introduced in Section 2.4 (Page 9 Lines 214−215): "** ***The 'nearest neighbor' approach was retained to match the grid point location from the ESA CCI SM product or CLM4.5 simulation with that of the in situ measurements*** **".**

2. Moreover, we added some discussions on the impact of the choice of soil depths for CLM4.5 on the statistical metrics against in situ observations and ESA CCI SM. Please see the discussion section (Page 20 Lines 478−498). The results are presented in Fig. 11. It is found that the choice of soil depth for CLM4.5 has a significant effect on the statics. The average R_{sp} against in situ observations ranged between 0.417 and 0.425 at different soil depths while the ESA CCI SM had higher R_{sp} values and a larger range, which was between 0.507 and 0.546. CLM4.5 agreed better with in situ observations at 0−10 cm than those at other soil depths, confirming that it is better to evaluate the CLM4.5 using the weighted average values at 0−10 cm. In contrast, the best agreement between the CLM4.5 simulation and ESA CCI SM product was found at the first layer (0−1.75 cm). These findings suggest that the mismatch in soil depths between the ESA CCI SM and in situ observations may have had a large effect on the statistical metrics, which is one of the reasons for its higher ubRMSD and lower R_{sp}.

3. It is noted that if more than one in situ station remained in a single 0.25° grid boxes, only one of them was included based on the correlation check method (Dorigo et al., 2015). The detailed description has been added to Section 2.3, as suggested. Please see Page 8 Lines 188−193.

5) The soil moisture observations of CMA were made by destructive sampling, which means that the sample each time is taken at a different location. Many studies (e.g. see work of Luca Brocca) have shown the enormous variability soil moisture can have even at local scales, in particular in absolute terms. The consequences of this sampling on your results needs to be thoroughly discussed.

Response: In response to this comment, we added a detailed description of in situ soil sampling method and a discussion of its potential effect on the comparison results to Section 4. Please see Pages 18−19 Lines 439−455. According to the user guide of in situ soil moisture measurements from agricultural meteorological stations of China Meteorological Administration (in Chinese, http://cdc.nmic.cn), soil moisture observation method is summarized as follows: (1) the observation field of each station is divided to four parts and four soil samples would be collected each time; (2) their soil moisture contents in dry weight basis (the ratio of water mass to dried soil's weight) are determined

by drying the soil, respectively; (3) the average value from the four samples are recorded as mass percentage for this station. It is noted that the horizontal distance between two successive samples at the same part of each station is no more than 2 meters, which leads to almost the same meteorological and soil conditions. Moreover, to reduce the effect of soil moisture heterogeneity, the station was usually chosen to be over flat surface. However, the original observations from four samples at each station were not available and we could not investigate the effect of destructive sampling on the comparison results in this study. Previous studies (Brocca et al., 2014a, b) showed that soil moisture had enormous variability even at local scales, in particular in absolute terms. These methods will be used in our future work to discuss the impacts of spatial variability and destructive sampling method.

6) It is not mentioned explicitly, but I understand that the precipitation measurements in ITP dataset used for forcing CLM include are taken at the same locations as the CMA sol moisture measurements. Therefore, it doesn't surprise me that the correlations obtained for the CLM simulations are higher than those obtained for soil moisture. To understand the real quality of your CLM simulations, the soil moisture fields should be validated at locations where no in situ precipitation measurements are made. You could do this by leaving the in-situ P measurements out, either entirely or by cross-validation (i.e. you use all in situ P measurements except the one made at the location of your soil moisture measurement).

Response: In response to this comment, we added a discussion about the effect of in situ precipitation measurements on the comparison results for CLM4.5 to Section 4. Please see Pages 19−20 Lines 462−477. It is noted that in situ precipitation measurements and soil moisture observations were obtained from different sources and were not collected at the same locations. However, the two types of data were still located in the same 0.25° grid box for many stations; it was necessary to investigate the effect of in situ precipitation measurements on the comparison results. The averaged statistical metrics over the sites with ("Rain") and without ("No Rain") in situ precipitation measurements are presented in new Fig. 10. It is noted that only the sites (200 of 306) with a significant ($p < 0.05$) positive Spearman's correlation for both the ESA CCI SM product and CLM4.5 simulation were considered, as described in Section 3.2. Figure 10 shows that, for the CLM4.5 simulation, in situ precipitation measurements (101 of 200) led to a slightly higher correlation (R_{sp} = 0.46) with in situ soil moisture measurements than was obtained without the precipitation measurements (R_{sp} = 0.44, 99 of 200). This may be related to the issues of spatial resolution. The observations from 740 operational stations of the CMA were merged with other meteorological forcing datasets to generate the ITP forcing data with a spatial resolution of 0.1° × 0.1°, while the CLM4.5 was run at 0.25° × 0.25° resolution in this study. If the CLM4.5 had been run at 0.1° × 0.1° resolution, the effect of ground-based precipitation on the comparison results may have been clearer (not shown in this study).

7) The selection of comparison metrics needs to be reconsidered: as stated above, for ESA CCI SM metrics of "absolute" deviation (i.e. bias, RMSD, SDV, E) basically provide an indication of how well GLDAS-Noah soil moisture fits the in situ observations. The metric E corrects for differences in mean soil moisture (additive bias), but not for differences in the dynamic range (multiplicative bias). As measure measure on its own, it is not very valuable; it mainly makes sense as part of the taylor diagram. The ubRMSD as computed from the anomalies does not correct for differences in variances between the in-situ data and the other data. It would make sense to compute the Spearman correlation coefficient in addition to the pearson R as a non-linear relationship between the in-situ data and the coarse scale products is expected because of differences in spatial support and depth (see Gruber et al. 2013, 2015 for valuable discussions on this issue.

Response: Based on this comment and the second comment, the statistical metrics used in the old manuscript, including the mean bias error (BIAS), root mean square difference (RMSD), normalized standard deviation (SDV), and centered normalized RMSD (E), have been removed; and we mainly focused on the analysis about the correlation coefficients and unbiased RMSD (ubRMSD) in the revised manuscript. Please see Section 2.4 (Pages 9−10, Lines 219−237). In addition, the Pearson correlation coefficient was instead by Spearman rank correlation (Dorigo et al., 2015), please see Page 10 Lines 235−237; both the ESA CCI SM and CLM4.5 soil moisture datasets were scaled into the dynamic range of the in situ observations using a linear rescaling method (Brocca et al., 2013; Dorigo et al., 2015) prior to computing the ubRSMD (new Eq. 2), please see Page 10 Lines 230−235. Old Fig. 8 (Taylor diagram) and Table 3 have been removed; other old figures and tables related to these metrics were also revised. Please see new Figs. 3−11 and Table 2.

8) Performances over 8 sub-regions: The regions are boxes which only roughly follow natural climate or land cover zones. Why did you define the regions in this way? And why are some regions not included in the analysis (e.g. the Tibetan plateau)? More importantly, the average performance statistics per region should be computed in a different way: first you should compute the statistics for the stations individually (you already do this), and then you average the numbers. It doesn't make sense to average in-situ observations that may have completely different soil moisture dynamics. You even mention this on p5163.l24-25

Response: Based on the comments, we added a detailed description of the definition of the eight sub-regions to Section 3.3. Please see Page 13 Lines 319−323. Furthermore, the analysis of average performance statistics using the method suggested by the reviewer was added to Section 3.3. Please see Pages 13−15 Lines 323−350. It is noted that several previous studies (Wang and Zeng, 2011; Liu and Xie, 2013) used the averaged soil moisture value of available in situ stations to represent the areal mean for the corresponding sub-region (only counting those grid cells closest to the relevant observation stations) and then computed their statistical metrics against in situ observations. Therefore, this method was also remained in the revised manuscript. Please see Page 15 Lines 351−363. In addition, because only four stations are available over the

Tibetan Plateau, the analysis over this region was not included in our study. In fact, the statistical metrics for the four sites can be observed from Figs. 3 and 4. For the ESA CCI SM product, the Spearman correlation ranges from 0.33 to 0.39 and the ubRMSD values are between 0.039 and 0.093 m^3 m^{-3}. For the CLM4.5 simulation, the Spearman correlation ranges from 0.40 to 0.54 (one site was not significant) and the ubRMSD values are between 0.040 and 0.103 m^3 m^{-3}.

9) There is some ambiguity regarding the terminology used: the metric E (centred RMSD) presented in Eq (5) is the same as the ubRMSD used in Albergel et al., 2013b, which again is the same as the RMSD of the anomalies (p5164.l26). So why do you present the same metric twice? In addition, the "real" ubRMSD should correct both for differences in the mean (additive bias) and the variance (multiplicative bias). Effort needs to be out in harmonising the metrics and terminology with existing literature.

Response: Based on the comment, the centered normalized RMSD (E) and the RMSD on soil moisture anomalies, called ubRMSD in the old manuscript (slide 5164 L.26), have been removed. Instead, only the ubRMSD (new Eq. 1) and Spearman rank correlation coefficient were analyzed in the revised manuscript. Moreover, a linear scaling method was used to scale both the ESA CCI SM and CLM4.5 soil moisture datasets into the dynamic range of in situ observations using a linear rescaling method based on the mean and standard deviation (Brocca et al., 2013; Dorigo et al., 2015). Please see Pages 9−10 Lines 219−237.

10) p5166.l18: you speculate that the introduction of the ASCAT soil moisture product in ESA CCI SM may have caused the decrease in skill for the period 2007-2011. But how can you then explain that the skill dramatically improves again for the latest blending periods, which just as well integrate ASCAT observations? Besides, Dorigo et al. (2015) showed that it was not the quality of the ASCAT product as such, but the way in which it was integrated into the blended product what explained the decrease in quality.

Response: Based on the comment, we revised the interpretation of the decrease in the skill of the ESA CCI SM for the period 2007−2011. Please see Page 17 Lines 420−423: "*The cause of this degradation is still not entirely clear, but it may be related to the resampling and scaling strategy used to incorporate a new active input product from ASCAT (Dorigo et al., 2015)*".

General:

1. The official name of the satellite-based product is ESA Climate Change Initiative soil moisture, or in brief ESA CCI SM. Please replace all occurrences of ECV-SM with the official name.

Response: All the occurrences of "ECV-SM" have been replaced by the official name "ESA CCI SM" in the revised manuscript, as suggested.

Minor:

1. Line 4: you cannot speak about THE land surface model simulation from CLM 4.5 as the model can be forced in many different ways. Your output is only one of many possible outputs.

Response: This sentence has been revised to be: "*soil moisture estimations from the Community Land Model 4.5 (CLM4.5), forced by observation-based atmospheric forcing data*", as suggested. Please see Page 1 Line 17−18.

2. p5155.l12: Later it becomes clear that you'll only use 301 out of the 778 available stations (p5162.l4), so the number 778 needs to be replaced

Response: Revised as suggested. After the quality control, 306 sites out of the 778 available stations were used in this study. Please see Page 5 Line 110.

3. What version of the ESA CCI SM product was used? I assume v02.0 or v02.1?

Response: The ESA CCI SM version 2.0 was used in this study: "*The ESA CCI SM version 2.0 (v2.0), released by the Vienna University of Technology in July 2014, was used in this study*". Please see Page 7 Lines 159−161.

4. p5157.l19: were the flags also applied?

Response: Based on the comment, we added an interpretation for the quality flags of the ESA CCI SM product: "*Quality "flags" of the input products were transferred to ESA CCI SM to mask pixels affected by snow coverage, temperature below 0 °C, dense vegetation, and pixels where the retrieval of soil moisture data failed (Dorigo et al., 2015)*". Please see Page 7 Lines 169−171. In addition, the data affected by snow cover and temperatures below 0 °C for both the in situ measurements and CLM4.5 predictions between March and October were masked using the quality flags of the ECV CCI SM product. Please see Page 9 Lines 205−208.

5. Section 2.4 Evaluation stretegy: details need to be given on how the data were assembled to monthly values: averages? Did you take into account the flags? Did you take into account differences in the number of observations per month? For each month, were the number of daily observations used to build the (supposed) mean values equal for both data sets? A valuable discussion on the importance of correct flagging and comparable temporal and spatial collocation of datasets was provided in Wagner et al., 2012, ISPRS

Response: Based on the comment, we added a detailed description about the data processes for the ECV SSI SM product and CLM4.5 simulation before their evaluation against in situ observations. Only daily data of the ESA CCI SM product and CLM4.5 simulation on the days of each month when in situ observations were available were used to compute their monthly mean values. Please see Page 9 Lines 211−213. However, the number of observations per month for the ESA CCI SM and CLM4.5 may be different due to many missing data of the ESA CCI SM product (Fig. 2). The differences in the number of observations per month were not taken into count in this study. In addition, for the ESA CCI SM product, quality "flags" of the input products were transferred to ESA

CCI SM to mask pixels affected by snow coverage, temperature below 0 °C, dense vegetation, and pixels where the retrieval of soil moisture data failed (Dorigo et al., 2015); the data affected by snow cover and temperatures below 0 °C for both the in situ measurements and CLM4.5 predictions between March and October were masked using the quality flags of the ECV CCI SM product. Please see Page 7 Lines 169−171 and Page 9 Lines 205−208.

6. p5159.l16: The relationship between E does not seem to be correct: Check Taylor (2001) to verify this.

Response: The statistical metric E has been removed as suggested.

7. p5164.l16-18: Also the ESA CCI SM is sensitive to precipitation (see e.g. Brocca et al., 2014), so this is not a valid explanation. the most plausible explanation would be the use of forcing data that was measured at the meteorological stations themselves.

Response: The explanation has been removed, as suggested.

8. p5164.l20-24: It is not clear what you mean with this statement: why would vegetation attenuation lead to stronger anomalies?

Response: The explanation has been removed, as suggested.

9. p5166.l17: the period of the 6th blending period should be Jan 2007-Sept 2011 (not 2007)

Response: Revised as suggested.

References

Brocca, L., Melone, F., Moramarco, T.,Wagner, W., and Albergel, C.: Chapter 17: Scaling and filtering approaches for the use of satellite soil moisture observations. In G. P. Petropoulos (Ed.), Remote sensing of land surface turbulent fluxes and soil surface moisture content: State of the art (pp. 562), Taylor & Francis, 2013.

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W., and Levizzani, V. : Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. J. Geophys. Res. Atmos., 119, 5128−5141, doi:10.1002/2014JD021489, 2014a.

Brocca, L., Zucco, G., Mittelbach, H., Moramarco, T., and Seneviratne, S. I.: Absolute versus temporal anomaly and percent of saturation soil moisture spatial variability for six networks worldwide, Water Resour. Res., 50, 5560–5576, doi:10.1002/2014WR015684, 2014b.

Dorigo, W. A., Gruber, A., De Jeu, R. A. M., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R. M., and Kidd, R.: Evaluation of the ESA CCI soil moisture product using ground-based observations, Remote Sens. Environ., 162, 380–395, 2015.

Liu, J. G., and Xie, Z. H.: Improving simulation of soil moisture in China using a multiple meteorological forcing ensemble approach, Hydrol. Earth Syst. Sci., 17, 3355–3369, doi:10.5194/hess-17-3355-2013, 2013.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, W. J., Subin, Z. M., Swenson, S. C., Thornton, P. E., Bozbiyik, A., Fisher, R., Kluzek, E., Lamarque, J.-F., Lawrence, P. J., Leung, L. R., Lipscomb, W., Muszala, S., Ricciuto, D. M., Sacks, W., Sun, Y., Tang, J., and Yang, Z.-L.: Technical description of version 4.5 of the community land model (CLM), NCAR Technical Note NCAR/TN-503+STR, National Center for Atmospheric Research, Boulder, CO, 420pp, 2013.

Shangguan, W., Dai, Y. J., Liu, B. Y., Ye, A. Z., and H. Yuan: A soil particle-size distribution dataset for regional land and climate modelling in China, Geoderma, 171–172, 85–91, 2012.

Wang, A., and Zeng, X.: Sensitivities of terrestrial water cycle simulations to the variations of precipitation and air temperature in China, J. Geophys. Res., 116, D02107, doi:10.1029/2010JD014659, 2011.