Evaluation of ORCHIDEE-MICT simulated soil moisture over China and impacts of different atmospheric forcing data

Zun Yin¹, Catherine Ottlé¹, Philippe Ciais¹, Matthieu Guimberteau¹², Xuhui Wang¹³⁴, Dan Zhu¹, Fabienne Maignan¹, Shushi Peng³, Shilong Piao³, Jan Polcher⁴, Feng Zhou³, Hyungjun Kim⁵, and other China-Trend-Stream project members

¹Laboratoire des Sciences du Climat et de l’Environnement, CNRS-CEA-UVSQ, Gif-sur-Yvette 91191, France
²UMR 7619 METIS, Sorbonne Universités, UPMC, CNRS, EPHE, 4 place Jussieu, Paris 75005, France
³Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China
⁴Laboratoire de Météorologie Dynamique, UPMC/CNRS, IPSL, Paris 75005, France
⁵Institute of Industrial Science, The University of Tokyo, Tokyo, Japan

* A full list of the China-Trend-Stream project members and their affiliations appears at the end of the paper.

Correspondence to: vyin@lsce.ipsl.fr

Abstract. Four atmospheric forcing datasets: GSWP3 (Global Soil Wetness Project Phase 3), PGF (Princeton Global meteorological Forcing), CRU-NCEP (Climatic Research Unit-National Center for Environmental Prediction) and WFDEI (WATCH Forcing Data methodology applied to ERA-Interim reanalysis data), are used to drive simulations in China by the land surface model ORCHIDEE-MICT. Simulated soil moisture is compared with in-situ and satellite datasets at different spatial and temporal scales in order to: 1) estimate the ability of ORCHIDEE-MICT (ORganizing Carbon and Hydrology in Dynamic Ecosystems: aMeliorated Interactions between Carbon and Temperature) to represent soil moisture dynamics in China; 2) demonstrate the most suitable forcing dataset for further hydrological studies in Yangtze and Yellow river basins; 3) understand the discrepancies of simulated soil moisture among simulations. Results showed that ORCHIDEE-MICT can simulate reasonable soil moisture dynamics in China (median \( r = 0.53 \); \( \text{RMSE} = 0.06 \text{m}^3\text{.m}^{-3} \)), but the quality varies with forcing data. Simulated soil moisture driven by GSWP3 and WFDEI shows the best performance according to \( \text{RMSE}_{\text{GSWP3}} = 0.05 \text{m}^3\text{.m}^{-3} \) and correlation coefficient \( r_{\text{WFDEI}} = 0.64 \) respectively, suggesting that both GSWP3 and WFDEI are good choices for further hydrological studies. The mismatch between simulated and observed soil moisture is mainly explained by squared bias (SB) and lack of correlation weighted by the standard deviation (LCS). Large SB suggests that the parameterization in ORCHIDEE-MICT should be calibrated for further study in China. High LCS and underestimated soil moisture in the North China Plain demonstrate possible significant impacts of human activities like irrigation on soil moisture variation, which was not considered in our simulations. Finally, the discrepancies (\( D \)) of meteorological variables and simulated soil moisture among the four simulations are analyzed. The result shows that the \( D \) of soil moisture is mainly caused by the \( D \) in precipitation frequency and air humidity rather than precipitation amount.
1 Introduction

Climate change significantly influences the hydrological cycle, which in turn affects ecosystems services, food security and water resources (Bonan, 2008; Piao et al., 2010; Seneviratne et al., 2010; Zhu et al., 2016). In China, annual precipitation increased in the South but declined in the North over the last several decades (Zhai et al., 2005; Ye et al., 2013). This dipole of precipitation trends is partly reflected in the discharge trends of Yangtze and Yellow rivers (Piao et al., 2010), but other factors than precipitation changes affect river discharge including changes in rainfall intensity, land surface state or condition and water management (Ayalew et al., 2014; Grillakis et al., 2016; Williams et al., 2015). A prerequisite to understand how precipitation changes translate into river discharge changes is to analyze and evaluate the various components of the surface water budget and especially the key variable relationships between precipitation and soil moisture (SM), result of the partition of precipitation between evapotranspiration, infiltration and runoff.

SM indeed plays a crucial role in adjusting local climate (Teuling et al., 2010; Seneviratne et al., 2013), regulating productivity and ecosystem dynamics (Schymanski et al., 2008; Yin et al., 2014) and affecting carbon budgets (Calvet et al., 2004). SM controls vegetation photosynthesis through transpiration, which in turn significantly influences surface temperature (Dai et al., 2004; Bonan, 2008). It also impacts the infiltration rate of precipitation water in the soil and its state before rainfall events determines the ratio of surface runoff to precipitation (Grillakis et al., 2016). Therefore SM is not only of importance in understanding land surface processes, but also is a key indicator for predicting and addressing extreme events, such as heat waves, floods and droughts (Teuling et al., 2010; Hirschi et al., 2011; Wanders et al., 2014).

To investigate the spatial and temporal SM variations, in-situ measurements (Robock et al., 2000; Liu et al., 2001; Piao et al., 2009; Dorigo et al., 2011) are too sparse and not always representative of larger scales. Although they can provide first hand records of SM fluctuations, the density of in-situ networks cannot meet the requirement for continental scale studies, and the different measurement techniques make it difficult to combine different datasets. Satellite-based SM products (Njoku et al., 2003; Su et al., 2003; Wagner et al., 2012; Dorigo et al., 2015) provide excellent spatial coverage and temporal sampling, but their accuracy varies between instruments and retrieval algorithms used (Liu et al., 2012). Moreover, these estimations concern only the first centimeters of soil but root zone SM cannot be directly assessed, unless a model simulating the water transfer processes is used.

Land surface models (LSMs) are able to simulate SM dynamics consistently with precipitation forcing and modeled runoff (Rebel et al., 2012; Xia et al., 2014; Pierdicca et al., 2015). The uncertainty of simulated SM depends on the accuracy of atmospheric forcing, in particular precipitation frequency and intensity and radiation. However LSMs complexity is a source of structural errors (missing processes) and biased parameters. Thus it is necessary to validate simulated SM by observations in order to diagnose the source of errors and estimate the ability of the chosen LSM to simulate SM dynamics in the area of interest.

In this study, we evaluated a set of SM simulations over China using the land surface model: ORCHIDEE-MICT (ORGanizing Carbon and Hydrology in Dynamic EcosystEms: aMeliorated Interactions between Carbon and Temperature) (Guimberteau et al., 2017), which integrates land surface hydrology, energy budgets and vegetation dynamics (prognostic phenology, photo-
synthesis and ecosystem carbon cycling) operating on a 30-min time step. Four atmospheric forcing data were used to drive the simulations. The resulting SM outputs were evaluated against different datasets including in-situ and spaceborne measurements in order to address three questions:

- Is the model able to provide a reasonable estimation of SM dynamics in China, as a prerequisite for simulation of river discharge in further studies?
- Which atmospheric forcing gives the best comparison with the available observations?
- Which meteorological factor explains the differences in SM among the simulations?

The atmospheric forcing and SM datasets used in this study are described in Section 2. Section 3 presents the model experiments. Evaluation of simulated SM and discussion are given in Section 4 and 5 respectively.

2 Forcing and evaluation datasets

2.1 Atmospheric forcing

Four gridded atmospheric forcing datasets are used to force the model over China: Global Soil Wetness Project Phase 3 (GSWP3), Global Meteorological Forcing dataset for land surface modeling (PGF), Climatic Research Unit–National Center for Environmental Prediction (CRU-NCEP) and the WATCH Forcing Data methodology applied to ERA-Interim reanalysis data (WFDEI). All input variables needed are the air temperature at 2 m \(T_t\), rainfall and snowfall rates, atmospheric specific humidity at 2 m \(Q_t\), surface pressure, downward short/long wave radiations \(R_s\) and \(R_l\) and wind speed \(W\). The four forcing datasets are combinations of reanalysis and observation data. These datasets, although built by different groups using different methods, are not independent from each other since they share some common input data. Detailed descriptions are listed below and general information is summarized in Table 1. Preprocessing of the datasets for ORCHIDEE-MICT is described in Sect. 3.3.

GSWP3

The Global Soil Wetness Project Phase 3 v0 (GSWP3, http://hydro.iis.u-tokyo.ac.jp/GSWP3) provides a 3-hourly climate data at 0.5° resolution from 1901 to 2010. It is based on the 20th Century Reanalysis (20CR; Compo et al. (2011)), which is downscaled from 2° to 0.5° by a spectral nudging technique in a Global Spectral Model (Yoshimura and Kanamitsu, 2008), in order to maintain both low and high frequency signals at high spatio-temporal scale. Single ensemble correction and vertically weighted damping are applied to remove known artifacts in high latitude regions (Hong and Chang, 2012; Yoshimura and Kanamitsu, 2013). Moreover, observation data are used for bias-correction, such as GPCC v6 (Global Precipitation Climatology Centre; Becker et al. (2013)) for precipitation, SRB (Surface Radiation Budget; Stephens et al. (2012)) for radiation and CRUTS v3.21 (Climate Research Unit; Harris et al. (2014)) for temperature.
PGF

The Princeton Global meteorological Forcing dataset for land surface modeling (PGF, http://hydrology.princeton.edu/data.pgf.php) provides 3-hourly data at 1° resolution from 1901 to 2012 (Sheffield et al., 2006). It is constructed by combining the NCEP-NCAR (National Centers for Environmental Prediction-National Center for Atmospheric Research) reanalysis of Kalnay et al. (1996) with several observation datasets. Precipitation is corrected by downscaled CRU TS v3.1, GPCP (Global Precipitation Climatology Project; Adler et al. (2003)) and TRMM (Tropical Rainfall Measuring Mission; Huffman et al. (2007)) data. SRB and CRU TS data are used in the assimilation of radiation and air temperature, respectively. Other variables (e.g., specific humidity, surface air pressure, wind speed) are just spatially downscaled from NCEP-NCAR according to the local elevation.

CRU-NCEP

The CRU-NCEP v6.1 (ftp://nacp.ornl.gov/synthesis/2009/frescati/modeldriver/cru ncep/analysis/readme.htm) provides 6-hourly 0.5° data. It combines the coarse temporal (monthly) resolution CRU TS dataset with the NCEP reanalysis, which has a higher time interval (6-hourly) but is only available at 2.5°. Monthly climate (except for precipitation) is identical to CRU TS, and NCEP is used only to reconstruct the 6-hourly variability within each month after bi-linearly interpolated to 0.5°. For precipitation the original NCEP values are used for temporal linear interpolation in those CRU grid cells (0.5°) covered by the specific NCEP grid cell (2.5°) in each month. CRU-NCEP dataset is available from 1901 to 2015 at global scale and it is updated each year with differences between versions.

WFDEI

The WFDEI forcing is generated by applying the WATCH Forcing Data methodology (http://www.eu-watch.org, Weedon et al. (2014)) to the ERA-Interim reanalysis (Dee et al., 2011) providing 3-hourly data at 0.5° from 1979 to 2009. The ERA-Interim blends GCM modeled variables and a suite of observations by a 4D-Var (4-dimensional variable analysis) data assimilation system (Weedon et al., 2014). All variables are bias-corrected using CRU TS. For precipitation, we use a version that has been bias-corrected with GPCC v5 and v6.

2.2 Soil moisture datasets

International Soil Moisture Network ISMN

ISMN is an international cooperative project providing a global gauged SM database (Dorigo et al., 2011) based on in-situ measurements from multiple monitoring regional sub-projects. Here only data from the CHINA sub-project is used (Robock et al., 2000) with in-situ volumetric water content (depth of water column over depth of soil in m^3.m^{-3}) from 40 stations between 1981 and 1999. SM profiles on 11 vertical layers were collected three times per month (on 8th, 18th and 28th of each month). The 11 sampled soil layers are: 0-5 cm, 5-10 cm, and then every 10 cm layers until 1 m. Most stations are located in
cropland or grasslands, but information about land use types and soil texture of each site are not provided. Moreover, there is no information about management practices affecting SM, such as irrigation or tillage.

In spite of the long length of this dataset, the data quality among stations varies widely and the monitoring period differs at each station. Some stations only recorded SM during the growing season, while others have a full year record. Furthermore, the measurements including the 5 deep layers (below 50 cm) are less than those including the top 6 layers. Only stations with more than 15 years of data were selected, which at least cover the same period (1984-1999). To make sure that there is at least half of the data available in the 15-year time series, stations with less than 270 measurement points in the top 6 layers are removed. This leads to selecting a subset of 20 stations, and given the sparseness of data below 50 cm, only SM in the top 6 layers is used for model evaluation.

**In-situ soil moisture from Peking University**

The SM was measured over over 778 stations of agro-meteorological stations over China by the Chinese Meteorological Administration (Xu, 2014) and collected and harmonized by the research team in Peking University (PKU; Piao et al. (2009)). The dataset provides 10-day SM variation during the growing season (mainly between May and September) from 1991 to 2007. It provides SM profiles in 7 soil layers (0-10 cm, 10-20 cm, 20-30 cm, 30-40 cm, 40-50 cm, 50-70 cm and 70-100 cm), but only the bottom four layers often have missing records. This dataset concerns exclusively croplands but there is no explicit information of soil texture and irrigation. Similar to the ISMN, the monitoring durations among gauging stations are different. Only 203 stations that cover the period of 1992-2006 are chosen.

**GLEAM v3.0A soil moisture**

The Global Land Evaporation Amsterdam Model (GLEAM) version 3.0 is a multiple algorithm, observation-based model reconstructing the components of the land evaporation process, including daily SM, evapotranspiration and interception at 0.5° resolution, with three sub versions described by Martens et al. (2017). Due to the short duration of version B (2003-2015) and C (2011-2015) only version A, covering the period 1980-2014, is used here. Note that radiation and air temperature used in GLEAM 3.0A are from ERA-Interim and precipitation is from MSWEP (Multi-Source Weighted-Ensemble Precipitation; Beck et al. (2016)).

Both surface and root-zone SM from GLEAM are used for model validation. The surface SM in the top 0-10 cm results from a combination of SMOS (the Soil Moisture Ocean Salinity satellite mission; Kerr et al. (2001)), ESA CCI SM (ESA Climate Change Initiative Soil Moisture; Liu et al. (2011, 2012); Wagner et al. (2012)) and the Community Noah land surface model SM fields in the GLDAS (Global Land Data Assimilation System; Rodell et al. (2004)) through a data assimilation system developed by Martens et al. (2016). Root-zone SM is derived with a 2-layer soil model based on mass balance. GLEAM provides SM in separate land-cover tiles of bare soil (0-10 cm), low vegetation (0-100 cm) and tall vegetation (0-250 cm). These tiles are based on MODIS Vegetation Continuous Fields (MOD44B; Hansen et al. (2003)).
ESA CCI soil moisture

The ESA CCI SM is a multi-satellite based product (Liu et al., 2011, 2012; Wagner et al., 2012). The daily SM is retrieved from a suite of microwave sensors spreading the period of 1979-2010 with 0.5° resolution. The representative soil layer depth is approximately 0.5-2 cm. Multi sensors ensure a long term records of SM dynamics, however the quality of the data varies with the change of available sensors and corresponding algorithms. Moreover, the remote sensing technique limits its ability to detect SM in frozen soils or under snow cover. Consequently, SM data is not available during winter in high latitude regions (e.g., Northern China). The data quality varies also along the period according to the number of instruments available and the increase of their temporal and spatial resolutions. The period after the launch of METOP-A-ASCAT (Advanced Scatterometer) at the end of 2006 appears indeed much more stable. Thus we chose the data between 2007-2009 for validation. Note that the lack of winter observations in Northern China influences the model evaluation metrics.

3 Land surface model, simulation protocol and model-data comparison metrics

3.1 Land surface model

ORCHIDEE (Organizing Carbon and Hydrology In Dynamic EcosystEms; Krinner et al. (2005)) is a physical-based land surface process model. It is mainly composed by two modules. The SECHIBA (surface-vegetation-atmosphere transfer scheme) module calculates the exchange of water and energy between land and atmosphere with a high time interval (half an hour). While the STOMATE (Saclay Toulouse Orsay Model for the Analysis of Terrestrial Ecosystems) module estimates the carbon cycle at daily time scale. The ORCHIDEE-MICT (aMeliorated Interactions between Carbon and Temperature, SVN version 3952; Zhu et al. (2015); Guimberteau et al. (2017)) is a recent version of ORCHIDEE including new processes as the interactions among frozen soil, snow, plants and soil carbon pools. It accounts for soil freezing, soil carbon discretization, snow processes and lateral water flows to improve the simulation of the main biogeochemical cycles in permafrost regions. It has been chosen in this study because China has a large permafrost area especially for the Tibetan Plateau, where originate both Yellow and Yangtze rivers (Jin et al., 2000). To simulate the soil water dynamics, ORCHIDEE-MICT (like the standard version of ORCHIDEE) uses a 11-soil layer scheme, whose depth increases exponentially until 2 m. The respective depths (in meters) of the calculation nodes are the following: 0.0005, 0.002, 0.006, 0.014, 0.03, 0.06, 0.12, 0.25, 0.5, 1.0, 1.75. Each grid cell is separated into three tiles, closely linked to the vegetation type above, for which the hydrologic budget is calculated separately. In this configuration, the 13 PFT (plant functional type) of ORCHIDEE-MICT are grouped in bare soil, trees and grass/crops to allow to better represent their specific hydrology. The amplitude of soil moisture depends on soil texture, which is a part of boundary conditions. Explicit description of the ORCHIDEE-MICT model can be found in Guimberteau et al. (2017).

There are two main outputs of SM in ORCHIDEE-MICT. The total SM (θt) indicates the total amount of soil water volume in the top 2 m soil layer in a grid cell. The SM profile (θp) records the vertical distribution of soil water content in the 11 soil layers. Note that the θp in each soil layer is an average value among the three soil tiles. The initial unit of ORCHIDEE SM is $m^3.m^{-3}$. 

3.2 Simulation protocol

Four simulations were performed driven by different forcing datasets described in Sect 2.1. In the simulations, CO$_2$ rise and land use change are taken into account but with no human processes like irrigation. The PFT map is from LUH2 (http://luh.umd.edu) and the soil texture map is from Zobler (1986). For the 3-soil texture scheme of Zobler86, the minimum residual and maximum saturated SM are 0.065 and 0.43 m$^3$ m$^{-3}$, respectively. The model domain covers the mainland of China ([85-124°E]×[20-44°N]). The spatial resolution is as same as the original atmospheric forcing (Table 1) and the simulation period covers 39 years, from 1971 to 2009, except for the one driven by WFDEI, which is from 1979 to 2009. To make sure that carbon (LAI and biomass) and water cycle variables can reach equilibrium, a 100-year spin-up was performed by repeating 10 times the forcing of the period 1971-1980 (for WFDEI, 50 times the period 1979-1980). Starting from the end of the spin-up, simulations were run from 1981 to 2009 and the output were re-gridded at 1°×1° to match the (coarsest) resolution of the output driven by PGF forcing.

The temporal resolution of forcing datasets is either 3-hourly or 6-hourly (CRU-NCEP), which is larger than the simulation time step of SECHIBA (30 min). To have a reasonable precipitation intensity and thus a good infiltration of water in the soil, the default precipitation splitting algorithm of ORCHIDEE is applied in our simulations. At the beginning of each forcing time step, if precipitation occurred, the precipitation amount (precipitation rate multiplied by the time interval of specific forcing) will be homogeneously distributed to the first half of the forcing time step.

3.3 Model-data comparison methodology and metrics

Comparison protocol

As the soil depths, periods and spatial-temporal resolutions are different in the four datasets (Sect. 2.2), we have to chose corresponding ORCHIDEE outputs for each comparison. To compare with the in-situ data of ISMN and PKU, we first extracted modeled daily SM profile ($\theta_p$) from the nearest grid cell for each station. Then the SM above a certain soil depth was chosen (50 cm for ISMN and 20 cm for PKU). PKU SM is provided in degree of saturation, defined as the volume ratio of actual water content to its maximum value when the soil is saturated. As the soil porosity is unknown, the PKU SM dataset cannot be directly compared with simulated SM from ORCHIDEE, which is defined from modeled porosity. To overcome this problem, normalization was applied on both datasets before comparison. The normalized data at each station and in the corresponding grid cell of the model is the ratio of the difference between the original value and its mean (during the observation period) to its standard deviation.

Regarding the definition of GLEAM SM (Sect. 2.2), we used the daily top 6-layer (approximately 9.2 cm depth) averaged SM and the total SM of ORCHIDEE to compare with GLEAM surface and root-zone SM, respectively. According to the sampled depth of the ESA CCI SM product, the daily top 4-layer (2.2 cm) averaged SM from ORCHIDEE is used. The period length and soil depth of each comparison are shown in Table 2.
Metrics

Pearson correlation coefficient (r) is calculated to estimate the correlation between simulated and observed SM. Daily SM corresponding to the measurement date reported in ISMN was collected to calculate r. As there is no date information from the 10-day PKU dataset, we used the 10-day averaged SM from ORCHIDEE for comparison.

The Root Mean Square Error (RMSE) is applied in order to estimate the temporal differences between simulation and observation. The same data pairs are used for RMSE calculation as the correlation coefficient except for PKU due to the normalization. Note that RMSE is related to the magnitude of SM, which varies significantly in China. To make it comparable in space, relative RMSE is calculated by dividing the mean of the simulated and observed SM.

According to Kobayashi and Salam (2000), the mean squared deviation (MSD), which is RMSE^2, can be decomposed into squared bias (SB), squared difference between standard deviation (SDSD) and lack of correlation weighted by the standard deviation (LCS), as:

\[
\text{MSD} = \text{RMSE}^2 = \text{SB} + \text{SDSD} + \text{LCS}. \tag{1}
\]

SB is the bias between simulations and observations. It is independent from other two components:

\[
\text{SB} = (\bar{s} - \bar{m})^2, \tag{2}
\]

where \(\bar{s}\) and \(\bar{m}\) are the mean of simulated and measured values, respectively. The SDSD indicates the mismatch of variation magnitude between simulated and observed variables, defined as:

\[
\text{SDSD} = (\text{SD}_s - \text{SD}_m)^2, \tag{3}
\]

where \(\text{SD}_s\) and \(\text{SD}_m\) are standard deviation of simulations and measurements, respectively. High SDSD implies a failure of the model in simulating the degree of fluctuation across the \(n\) measurements. Note that SDSD correlates with LCS, which accounts for \(\text{SD}_s\) and \(\text{SD}_m\) as well:

\[
\text{LCS} = 2\text{SD}_s\text{SD}_m(1 - r), \tag{4}
\]

where \(r\) is the Pearson correlation coefficient. The LCS is an indicator of the performance of the model to simulate the pattern of fluctuation of the measurements. The lower the LCS is, the better the model performs.

Finally, to evaluate the characteristic time scale of modeled SM response to hydrological processes, lag-\(k\) autocorrelation coefficient (\(R_k\)) is computed. The \(R_k\) is the correlation coefficient of a time series with itself but with a \(k\) time step lag, as:

\[
R_k = \frac{\sum_{i=1}^{n-k}(x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{5}
\]

where \(n\) (\(n > k\)) is the length of the specific time series; \(\bar{x}\) is the mean value. For SM time series in a specific grid cell, \(R_k\) was computed for different \(k\) values. The value of \(R_k\) decreases with increasing \(k\) and the \(k\)-lag time series are considered not auto-correlated if \(R_k\) is less than a threshold \(1/e\) (Maurer et al., 2001; Rebel et al., 2012). The day number when \(R_k\) first drops...
below a threshold of $1/e$ is called number of lag days (NLD). The NLD difference is used to compare the overall characteristic time scales between datasets. The difference of $R_k$ profiles gives additional information on the autocorrelations for lag. The $R_k$ comparison was implemented between GLEAM and ORCHIDEE because other datasets do not have complete daily records.

The linear trend of SM change in the 29 years is of interest as well. The Mann-Kendall test (Mann, 1945; Kendall, 1975) is applied to test if simulations capture observed trends of SM, with $p$-value $< 0.05$ indicating a significant trend.

### 3.4 Correlation of uncertainties between SM and meteorological factors

In our simulations, the difference in atmospheric forcing is the only source of difference in simulated SM. We look at different indicators related to precipitation to explain SM differences among simulations. These indicators are monthly precipitation amount ($P$) and the number of precipitation days in one month ($N_P$) excluding days with $P < 0.01$ mm.d$^{-1}$. Precipitation days are categorized into 5 classes of 0.01-1, 1-5, 5-10, 10-15 and $> 15$ mm.d$^{-1}$ and the number of days with precipitation amount in each class was calculated, denoted by $N^i_P$ with $1 \leq i \leq 5$. Monthly values of other meteorologic variables (incoming short/long wave radiation, air temperature, air humidity and wind speed) are also considered in the analysis. Regarding SM, both total SM ($\theta_t$) and SM in each soil layer ($\theta_{t,i}$, $i$ is the index of soil layer) were correlated with these indicators. To estimate the difference of a variable $x$ among the four simulations, the averaged MSD ($D_x$) is computed as:

$$D_x = \frac{1}{n} \sum_{i \neq j} \sum_{t=1}^{n} (x_{t,i} - x_{t,j})^2 \binom{n}{2},$$  \hspace{1cm} (6)

where $N = 4$ is the number of simulations; $i$ and $j$ ($1 \leq i, j \leq N$) are indexes of the four simulations; $\binom{n}{2}$ is the binomial coefficient; $n$ is the length of the time series; $t$ is the time step. Note that we use the absolute value of $D_x$ not relative $D_x$ ($D_x$ over averaged value of $x$ in the specific grid cell) for the analysis because the relative $D_x$ cannot reflect the linkage of uncertainty between inputs and outputs. Detailed explanation is shown in Supplement B.

### 4 Results

#### 4.1 Spatial patterns of precipitation and simulated soil moisture

China contains multiple climate regimes, which makes the hydrological situations of Yellow and Yangtze Rivers notably complicated. The Yellow River basin (YLRB, magenta contour in Fig. 1) is mainly located in semi-arid and temperate zone (Kottek et al., 2006), whereas the Yangtze River basin (YZRB, red contour in Fig. 1) is in a humid subtropical region. The climatic difference between the two river basins is reflected from the seasonality of $P$ and $\theta_t$ (Fig. A1). The maximum monthly $P$ in YZRB (175 mm.mon$^{-1}$) occurs in June, which is one month earlier than that in YLRB (105 mm.mon$^{-1}$). As a consequence, the $\theta_t$ of YZRB starts to increase after May, while the increasing trend is one month delay in YLRB.
The 29-year mean annual precipitation (MAP) and simulated total SM ($\theta_t$) are compared for the four forcing data (Fig. 2). The MAP in China is strongly influenced by both Western North Pacific Monsoon and Indian Monsoon, resulting in a decreasing trend from the southeast (2450 mm yr$^{-1}$) to northwest (9 mm yr$^{-1}$) parts of China. In spite of similar spatial gradients, the $\theta_t$ patterns do not perfectly coincide with that of the MAP. For instance, $\theta_t$ with GSWP3 in the North China Plain (Fig. 1) is larger than that with other forcing datasets, although MAP is similar. Another mismatch occurs in East Tibet ([93-99°E]×[27-39°N]), where higher SM is found in PGF and WFDEI but MAP has no much difference among the forcing datasets.

4.2 Soil moisture evaluation against multiple datasets

Comparison with ISMN and PKU in-situ data

In most cases, the correlation between modeled and measured SM at selected ISMN stations (see Sect. 2.2) is significantly positive for $r$ value (Fig. 3). High correlations ($r > 0.6$) are found over the Loess Plateau in the semi-arid zone. In the North China Plain, $r$ is much lower (between 0.2 and 0.5). Three ISMN stations (marked by squares in Fig. 3(a)) are chosen to represent for different wet conditions and model-data comparisons are shown in Figure 4. Xifeng station located in the semi-arid zone (MAP = 556 mm yr$^{-1}$), where $\theta_t$ is low (0.2 m$^3$.m$^{-3}$ on average) with a large inter-annual variation. The variability of simulated $\theta_t$ is consistent with observations (0.55 < $r$ < 0.87; when CRU-NCEP is excluded) due to lower human impacts on rainfed agriculture in this region (Li et al., 2014). Xinxian station is located in the North China Plain with similar MAP (580 mm yr$^{-1}$) than at Xifeng, but in a traditional irrigation region (Wang et al., 2016). $\theta_t$ is underestimated, possibly because irrigation is not included in our simulations. Thus, the model cannot capture the seasonal variations of $\theta_t$, given $r$ values ranging between 0.11 and 0.23. Xuzhou station is in the North China Plain as well but with a higher MAP (847 mm yr$^{-1}$). The fluctuation of simulated and observed $\theta_t$ are coherent, leading to $r$ from 0.55 to 0.66. However, the magnitude of $\theta_t$ is systematically underestimated as well.

The correlation coefficients of $\theta_t$ between simulations and PKU dataset are shown in Figure 3 as circles. Modeled $\theta_t$ has a better performance in the Loess Plateau and North China Plain than other regions, suggesting that ORCHIDEE is better able to capture the variations of SM in semi-arid and temperate zones. In comparison to ISMN, $r$ between simulated $\theta_t$ and PKU observations is lower. This may be caused by the shallower depth of the PKU data (20 cm) with stronger influence from fast infiltration and transpiration processes than in the ISMN records (1 m). Moreover, the PKU dataset only records $\theta_t$ during the growing season, leading to lower $r$ in absence of full seasonal variations.

Comparison with GLEAM v3.0A observation-based data product

The left panel of Figure 5 shows correlation coefficients between GLEAM surface SM ($\theta_s$) and corresponding modeled SM in the surface layer (0-10 cm). Simulated $\theta_s$ is significantly correlated with GLEAM (median $r = 0.57$) except for a part of Northwestern China ($r < 0.3$). At the middle stream of YZRB, $r$ is lower than its surroundings. According to the spatially averaged $r$ of $\theta_s$, GSWP3 (0.58) and WFDEI (0.69) lead to better performances with ORCHIDEE than the PGF (0.46) and CRU-NCEP (0.54). Note that both WFDEI and GLEAM v3.0A used ERA-Interm reanalysis to reconstruct the time series of
precipitation, which can explain the higher $r$ when ORCHIDEE is forced by WFDEI. The correlation coefficients of simulated and GLEAM root-zone SM ($\theta_s$, Fig. A2) have the same patterns than for $\theta_s$ but higher values (median $r = 0.64$) due to the lower variability of $\theta_s$, which smoothes out misfits related to differences in individual rainfall events between ORCHIDEE and GLEAM for $\theta_s$.

Figure A3 and 5 (right panel) show the relative RMSE and the MSD decomposition of $\theta_s$ between GLEAM and ORCHIDEE. Low relative RMSE ($< 0.3$) is found in most regions except for the North China Plain ($> 0.5$), where the RMSE is dominated by the squared bias values (SB) between simulation and observation as shown in Fig. 5(b). This is clearly shown in the relative difference (Fig A4) between GLEAM and ORCHIDEE where simulated $\theta_s$ is approximately 30% lower than in GLEAM. Southern China has lower relative RMSE ($< 0.2$), and MSD is dominated by SB as well. Different from the North China Plain, SB in southern China may be due to the mismatch of land cover and soil parameterization between ORCHIDEE and GLEAM. For instance, the saturated SM in South China is 0.36 m$^3$.m$^{-3}$ while the maximum SM in GLEAM is 0.45 m$^3$.m$^{-3}$. A high contribution of LCS to MSD is found in Qinghai-Tibetan Plateau, the upper part of YZRB and YLRB, suggesting a mismatch of the phase of SM variability. The MSD is dominated by SDSD in northwestern China ($P < 200$ mm.yr$^{-1}$) suggesting different magnitudes of SM fluctuations. Nevertheless, the RMSE in Qinghai-Tibetan Plateau and northwestern China is as low as in southern China ($< 20\%$). Overall, ORCHIDEE is able to give a reasonable estimation of $\theta_s$ in regions where irrigation is not widespread.

Figure 6(a)-(e) shows NLD of ORCHIDEE and GLEAM $\theta_s$ computed based on the $k$-lag autocorrelation coefficient $R_k$. High NLD implies that $\theta_s$ has a longer memory in response to rainfall inputs. However the spatial distribution of NLD depends not only on rainfall frequency and intensity but also on evapotranspiration and runoff losses after SM recharge by rainfall. The NLD patterns of GLEAM and ORCHIDEE $\theta_s$ are similar, which is encouraging in terms of how ORCHIDEE simulates the processes controlling the decrease of SM after each rainfall. Both southern and southeastern China have higher NLD, like in GLEAM. Lower NLD ($\approx 1$ day) prevail around 30°E in eastern China, whilst the North China Plain has NLD values of 40 days. The main difference of NLD between GLEAM and ORCHIDEE is in Inner Mongolia and over the Loess Plateau, where the ORCHIDEE $\theta_s$ has values of 10 days, against 35 days in GLEAM.

$R_k$ of spatially averaged $\theta_s$ in three regions is shown in Figure 6(f)-(h). Overall, $R_k$ of ORCHIDEE $\theta_s$ is consistent with that of GLEAM. In all China and YLRB, $R_k$ of GLEAM is larger than that of ORCHIDEE, suggesting that modeled $\theta_s$ has a faster response to rainfall input. In YLRB, the difference is larger, where $R_k$ of ORCHIDEE and GLEAM $\theta_s$ are significant until 30 days and 41 days, respectively. ORCHIDEE $R_k$ in the YZRB is higher than GLEAM $R_k$ when $k < 14$. Moreover, the $R_k$ of GLEAM and ORCHIDEE is very close except for the case of GSWP3. The bias of $R_k$ can be explained by higher/lower simulated evapotranspiration in YLRB/YZRB in ORCHIDEE than that in GLEAM (not shown), through which the decline of ORCHIDEE $\theta_s$ is faster/slower after rainfall events and lead to a lower/higher $R_k$.

The trend of ORCHIDEE $\theta_s$ (Fig. A5) is similar to that of $P$ (Fig. A6), which is less significant than that of GLEAM $\theta_s$ (Fig. A5). In northwestern China, increasing $\theta_s$ is found in simulations ($< 0.2 \times 10^{-3}$m$^3$.m$^{-3}$.yr$^{-1}$) and GLEAM ($0.2-0.4 \times 10^{-3}$m$^3$.m$^{-3}$.yr$^{-1}$). The trend may be due to increasing $P$ (Fig. A6). GLEAM $\theta_s$ decreased dramatically in eastern China ([103-122°E]×[20-35°N]) while the trends of ORCHIDEE $\theta_s$ are not homogeneous in this region. In addition, all forcing
datasets show an increasing $P$ in the North China Plain, which leads to slight increase of simulated $\theta_s$. But GLEAM shows decreasing $\theta_s$ in most area of the North China Plain. The mismatch of the $\theta_s$ trend suggests that climate change (change of precipitation regimes) is not enough to predict the trend of SM.

**ESA CCI SM**

Figure 7 (left panel) shows $r$ between ESA CCI and ORCHIDEE $\theta_s$. In the North China Plain, $r$ is above 0.5 except for PGF. Lowest $r$ is found in the downstream of YZRB. In southern China, simulated $\theta_s$ is better correlated with the ESA CCI $\theta_s$ (e.g. in Yunnan province $r \approx 0.8$). Weak correlations ($r < 0.3$) only exists in the transition zone from the south to the north. Compared to GLEAM $\theta_s$, the $r$ between ORCHIDEE and ESA CCI $\theta_s$ is much lower. It is due to a shallower depth of the ESA CCI $\theta_s$, which is more sensitive to surface process errors and difficult to simulate with a LSM. Moreover, ESA CCI $\theta_s$ is a purely satellite product while GLEAM $\theta_s$ (v3.0A) is a combination of modeled, in-situ and satellite SM. The latter one is not totally independent of the forcing datasets and therefore more comparable to our simulations.

However, the relative RMSE (0.32 in spatial average, Fig. A7) is lower than that in the GLEAM case (0.42). The components of MSD between ORCHIDEE and ESA CCI $\theta_s$ are shown in the right panel of Figure 7. The main source of MSD is LCS (phase mismatch). It implies that the magnitude of the simulated $\theta_s$ is reasonable, but the timing of the fluctuations differs between ORCHIDEE and ESA CCI. It is reflected from the relative difference between ORCHIDEE and ESA CCI $\theta_s$ (Fig. A8), the absolute values of which in 70% grid cells are less than 0.1. Moreover, the patterns of relative difference suggests that $\theta_s$ is underestimated in north China especially in YLRB.

**4.3 Comparison of the four forcing datasets**

The annual value of meteorological variables averaged over China of the four forcing datasets are compared in different regions (Fig. A9). Except for air temperature, differences can be found in all variables. For precipitation, air humidity, longwave and shortwave radiation, there are good agreements among GSWP3, PGF and WFDEI, however CRU-NCEP shows obviously differences from others.

To find the most realistic forcing dataset for SM performance given the ORCHIDEE model, several metrics were calculated and shown in Figure 8. Radar charts show the correlation coefficients ($r$) and RMSE of simulated SM in comparison to different datasets. Histograms show MSD and its three components. The median of specific metrics is listed in Table 3. GSWP3 has the best performance in estimating the magnitude of SM (lowest MSD) while WFDEI shows the best score in simulating SM variation (highest $r$). PGF provides as good estimation as GSWP3 in the YZRB, but performs more poorly in capturing SM variation in YLRB, which is also reflected from the components of MSD. The largest MSD is found in CRU-NCEP in most of comparisons, which is mainly contributed by SB. The SDSD and LCS of CRU-NCEP are also larger than others but the differences are not as significant as SB.

Thus we conclude that both GSWP3 and WFDEI are suitable to simulate SM dynamics in China with ORCHIDEE. The best choice can be made based on the main focus of specific research. For estimating magnitude of SM, GSWP3 is preferable;
for investigating SM variation, WFDEI is the best choice. Note that this study only provides the evaluation of SM, but other hydrological components should be compared with observations to confirm the superiority of GSWP3 and WFDEI.

4.4 Source of soil moisture difference among simulations

By investigating the $D$ of meteorological variables and simulated SM among the four simulations ($D_x$ for variable $x$; Eq. 6), two questions are addressed: (1) How is $D$ of simulated SM and forcing variables spatially distributed? (2) Can spatial patterns of $D$ of SM be explained by that of meteorological variables? Note that the relative value of $D$, $D$ over the magnitude of specific variable in each grid cell, is not suitable for the analysis (detailed explanation is in Supplement B). In short, the $D$ of input (meteorological variables) varies with the input magnitude. However the $D$ of output (simulated SM) depends on the input magnitude as well, not on the magnitude of output. This judgment is also confirmed by Figure A10, which shows that the $D$ of most inputs (except incoming shortwave radiation) increases with increasing input magnitude but the $D$ of simulated SM does not depend on its magnitude. Thus the relative $D$ weakens the linkage between uncertainties of inputs and outputs and is not proper to use here.

Figure 9 shows maps of $D$ of $\theta_t$ and meteorological variables. As the unit of $D$ depends on specific variables, it can only be used to compare spatial distributions, not values. High $D_{\theta_t}$ is found in southwest of China ([92-104°E]×[28-35°N]). However, similar patterns do not exist in the $D_P$ (Fig. 9(b)), suggesting that the difference of simulated $\theta_t$ is not caused by the difference of precipitation amount of forcing data. Similarly, in southwestern China, no high $D$ is found in meteorological variables except for number of precipitation days ($N_p$) and air humidity ($Q_a$), although the patterns of $D_{Q_a}$ overlap with $D_{\theta_t}$ but extend to zones with low $D_{\theta_t}$ as well (Fig. 9(c) and (g)).

To look for clearer links between input and SM $D$, we decompose $N_p$ and $\theta_t$ by scales of $P$ and soil layers respectively. Figure A11 and A12 show the $D$ of SM profile ($\theta_p$) and $N_p$ with different precipitation intensities as a function of MAP, respectively. The $D_{\theta_p}$ does not rise with increasing MAP, confirming that the magnitude of $D_{\theta_p}$ is not linearly correlated to the wetness conditions. Correlations between decomposed $N_p$ and $\theta_t$, however are more visible. Similar patterns are indeed found in Fig. A12(a) and (b), implying a strong linkage between the variation of simulated SM and the differences of slight precipitation (0.01 $< P < 5$ mm.d$^{-1}$) events among the forcing datasets.

The $r$ of $D$ between simulated SM and meteorological variables are shown in Figure 10. $D_{Q_a}$ and $D_{N_p}$ are highly correlated with $D_{\theta_t}$, implying that the difference of simulated $\theta_t$ can be explained by the differences of $Q_a$ and $N_p^2$ among the four forcing datasets. The $r$ between $D_{\theta_t}$ and $D_P$ is less than 0.4. All in all, these results suggest that the temporal distribution of precipitation is more important than its amount in influencing SM dynamics differences among the responses to the forcings.

Regarding $D_{\theta_p}$, highest $r$ ($=0.7$) is found with $N_p$, $N_p^2$ and $N_p^3$, which is consistent with $D$ patterns shown in Fig. A11 and A12. This indicates that SM difference in the first soil layer (0-1 mm) is mainly explained by the differences in light precipitation events among forcing datasets. However, the $r$ between $D_{N_p}$ and $D_{\theta_p}$ decreases with increasing layer depth. In the second layer (1-4 mm), $r$ is less than 0.5. On the other hand, $D_{N_p}$ keeps a high correlation with $D_{\theta_p}$ in deeper soil layers. The value of $r$ between $D_{N_p}$ and $D_{\theta_p}$ is approximately 0.8 from the second to the sixth layer (depth = 9.2 cm), and then drops smoothly until 0.5. The $r$ between $D_{\theta_p}$ and $D_{N_p}^2$ ($i > 2$) is lower, suggesting that the impact of $N_p$ differences on SM
differences decreases with increasing soil depth. Moreover, precipitation events with $1 < P < 5 \text{ mm.d}^{-1}$ is more important than others in affecting $D$ of SM. High $r$ values are also found between $D_{Q_a}$ and $D_{ip}$.

5 Discussion

5.1 Performance of the model to simulate soil moisture

Due to the spatio-temporal complexity of SM and its vertical profile, four datasets were selected to drive the simulations and modeled SM at different depths was validated against multiple datasets. The results showed that ORCHIDEE SM coincides well with GLEAM (median $r = 0.57$; median relative RMSE = 0.37) and ESA CCI SM (median $r = 0.45$; median relative RMSE = 0.31) in comparison to other model studies (Lai et al., 2016).

Higher $r$ were systematically found in southern China, the Loess Plateau and North China Plain; lower $r$ were found in northwestern China, western Tibetan Plateau, eastern Sichuan basin and downstream of YZRB. SM underestimation is significant in the Loess Plateau and North China Plain with modeled values being 30% and 20% less than in GLEAM and ESA CCI respectively (Fig. A4 and A8). It is not only due to model parameterization but also due to irrigation activities in those agricultural regions (Fig. 1), which are not considered in the simulations.

Because the in-situ SM measurements were collected only for croplands and grasslands (Robock et al., 2000; Piao et al., 2009), implying potential disturbances from human activities, $r$ was low in the comparison to ISMN and PKU datasets (median $r = 0.37$, Fig. 3). For instance, drought occurred in northern China during 1987-1988 (Trenberth et al., 1988; Yang et al., 2012), which is reflected in the variation of measured SM at Xifeng station (Fig. 4(a)). ORCHIDEE successfully reproduced the drought induced decline at Xifeng and Xinxian (Fig. 4(a)-(b)). But SM measured at Xinxian was maintained at a high level. A possible explanation is that the soil at Xinxian was irrigated. Consequently SM at Xinxian did not vary with precipitation leading to a low $r (< 0.23)$.

From the results we conclude that the version of ORCHIDEE used in this study provides a satisfactory simulation of SM dynamics in China, except in areas subject to irrigation. This calls for inclusion of irrigation and realistic crop phenology (Wang et al., 2017) as a priority for future application of this model for SM and river discharge dynamics.

5.2 Linkage of discrepancies between meteorological factors and soil moisture through ORCHIDEE

In Section 4.4, we showed that the spatial differences of simulated SM among the four forcing datasets were highly correlated with forcing differences in $N_p$ and $Q_a$. This suggests that precipitation frequency is more critical than precipitation amount in determining SM patterns, as pointed out by other studies, especially in arid and semi-arid regions (Baudena and Provenzale, 2008; Piao et al., 2009; Cissé et al., 2016). To precise the result, we studied the correlation coefficients between the spatial averaged $D$ of SM in different soil layers and of $N_p$ categorized by classes of precipitation intensity (Fig. 10). The result showed that differences in small rainfall events $N'_p$ with $1 < P < 5 \text{ mm.d}^{-1}$ are more important than other precipitation classes in explaining SM differences due to atmospheric forcing datasets.
Differences in $Q_a$ were also shown to explain a large fraction of the simulated SM differences across different forcings. $Q_a$ determines vapor pressure deficit, which in turn controls transpiration (Farquhar and Sharkey, 1982) and evaporation (Monteith, 1965), suggesting a strong control by atmospheric dryness of the differences in SM found among the four forcing datasets. Both of $Q_a$ and $N_p$ have positive impacts on SM, which enhances the correlation in Figure 10.

Estimating impacts of meteorological factors on SM dynamics is difficult. First of all, the importance of a meteorological variable on SM may vary with climate regimes. For instance, the importance of precipitation and radiation on SM changes from water to energy limited regions. Secondly, impacts of meteorological variables can be nonlinear through interactions with local ecosystem (Seneviratne et al., 2010), suggesting that even with same meteorological variable the simulated SM can be totally different (e.g., with different soil texture or land cover types). Moreover, SM can be strongly coupled with atmosphere (Koster, 2004; Taylor et al., 2012), implying that meteorological factors can be influenced by SM as well (such as cloudiness, precipitation, air humidity, etc), which is not included in this study. However, the logic of our importance analysis is simple. If the model inputs (forcing data) were the same, the outputs (SM) should be the same. In other words, the differences of outputs can only be caused by the difference of inputs in our simulation results. It does not matter whether the quality of atmospheric forcing is good. On the contrary, the more differences exists among these forcing datasets, the better our analysis is. To keep the analysis simple, we did not investigate temporal correlations in each pixel but focused on spatial patterns of $D$ at continental scale. Therefore, our results provided a general estimation of the importance of meteorological variable uncertainties to SM simulation through ORCHIDEE.

Indeed this approach is not able to demonstrate explicit links between meteorological variables and SM. We underlined the impacts of $N_p$ and $Q_a$ uncertainties, but it does not mean that other factors are unimportant. For instance, assuming that a variable can strongly influence simulated SM, if there was no much difference of the variable among forcing datasets, its importance cannot be detected in this work. Moreover, only one model was used in this study. Although ORCHIDEE performed very well in SM simulation, the lack of unknown mechanisms may weaken the linkage between SM and specific atmospheric variables. In one word, our analysis only focused on the inputs and outputs of the model and tried to diagnose the relationship between their differences.

6 Conclusions

Simulations in China were performed in ORCHIDEE-MICT driven by different forcing datasets: GSWP3, PGF, CRU-NCEP and WFDEI. Simulated soil moisture was compared to several datasets to evaluate the ability of ORCHIDEE-MICT in reproducing soil moisture dynamics in China. Results showed that ORCHIDEE soil moisture coincided well with other datasets in wet areas and in non irrigated areas. It suggested that the ORCHIDEE-MICT was suitable for further hydrological studies in China. However, the abnormal variation of observed SM in North China Plain implied potential impacts of irrigation, which was recommended to be considered in further simulations. Moreover, results showed that bias was mainly from model parameterization and atmospheric forcing. Thus parameterizations in ORCHIDEE-MICT should be calibrated and atmospheric forcing should be carefully selected to meet the situation of China.
Several criteria were chosen and compared among the four simulations in China, YZRB and YLRB. Results showed that GSWP3 and WFDEI, which had the best performances in correlation coefficients and RMSE respectively, are ideal choices for hydrological study in China. However, higher MSD in the Yellow River basin reflected the complicated climate condition in northern China, which might be significantly influenced by human activities as well. Finally, we used the difference of simulated soil moisture and meteorological variables to simply investigate the linkage between them. Results showed that the differences of simulated soil moisture were mainly explained by the differences of air humidity and precipitation frequency among the four forcing tested. However this coarse analysis cannot give explicit explanations about related mechanisms. Further study is needed to discover the interactions between soil water and climate through tracing the surface hydrological cycles and energy balances.

**Code and data availability.** The SVN version of ORCHIDEE-MICT used in this study is 3952, which is available at https://forge.ipsl.jussieu.fr/orchidee/wiki/DevelopmentActivities/ORCHIDEE-MICT-IMBALANCE-P. The ORCHIDEE code and scripts of analysis are available by contacting the correspond author. The GSWP3, PGF, WFDEI, ISMN, GLEAM and ESA CCI datasets are freely available online; for the CRU-NCEP and PKU datasets, please contact the corresponding author and Prof. Shilong Piao (slpiao@pku.edu.cn) respectively.

Members of the China-Trend-Stream project are (alphabetically): Catherine Ottlé¹, Feng Zhou², Hui Yang², Jan Polcher³, Laurent Li³, Matthieu Guimberteau¹,⁴, Patrice Dumas⁵, Pengfei Shi⁶, Philippe Ciais³, Shilong Piao², Shuai Wang⁷, Shushi Peng², Tao Yang⁶, Xiaoming Feng⁷, Xuanze Zhang², Xudong Zhou⁶, Xuhui Wang¹,²,³, Yi Xi² and Zun Yin¹.

¹Laboratoire des Sciences du Climat et de l’Environnement, CNRS-CEA-UVSQ, Gif-sur-Yvette 91191, France
²Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China
³Laboratoire de Météorologie Dynamique, UPMC/CNRS, IPSL, Paris 75005, France
⁴UMR 7619 METIS, Sorbonne Universités, UPMC, CNRS, EPHE, 4 place Jussieu, Paris 75005, France
⁵Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Avenue Agropolis, 34398 Montpellier Cedex 5, France
⁶State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Center for Global Change and Water Cycle, Hohai University, Nanjing 210098, China
⁷State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

**Author contributions.** P. Ciais, C. Ottlé and Z. Yin designed research. Z. Yin performed research, analyzed data and wrote the draft; all authors contributed to interpreting results, discussing findings and improving the manuscript.
Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This study was supported by the National Natural Science Foundation of China (grant number 41561134016) and by the CHINA-TREND-STREAM French national project (ANR Grant No. ANR-15-CE01-00L1-0L). M. Guimberteau and P. Ciais acknowledge support from the European Research Council Synergy grant ERC-2013-SyG-610028 IMBALANCE-P. We thank Brecht Martens and Suxia Liu for helpful discussion about GLEAM and ISMN datasets.
5 References


Tables

**Table 1.** General information of the climate forcing datasets. ‘Reanalysis’ and ‘Observations’ are corresponding datasets used in producing the atmospheric forcing. Detailed description can be found in Sect. 2.1.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Resolution</th>
<th>Duration</th>
<th>Reanalysis</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spatial</td>
<td>Temporal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSWP3</td>
<td>0.5°</td>
<td>3-hourly</td>
<td>1901–2010</td>
<td>20CR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GPCC, CRU TS, SRB</td>
</tr>
<tr>
<td>PGF</td>
<td>1°</td>
<td>3-hourly</td>
<td>1901–2012</td>
<td>NCEP-NCAR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CRU TS, GPCP, TRMM, SRB</td>
</tr>
<tr>
<td>CRU-NCEP</td>
<td>0.5°</td>
<td>6-hourly</td>
<td>1901–2015</td>
<td>NCEP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CRU TS</td>
</tr>
<tr>
<td>WFDEI</td>
<td>0.5°</td>
<td>3-hourly</td>
<td>1979–2009</td>
<td>ERA-Interim</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CRU TS, GPCC</td>
</tr>
</tbody>
</table>

**Table 2.** Summary of the soil moisture datasets for validation. ‘M+RS+RA’ indicates that the dataset is a model output driven by both remote sensing and reanalysis data. More details can be found in Sect. 2.2.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Type</th>
<th>Unit</th>
<th>Resolution</th>
<th>Duration</th>
<th>Contents</th>
<th>Corresponding ORCHIDEE soil layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISMN</td>
<td>in-situ</td>
<td>m$^3$.m$^{-3}$</td>
<td>station, 10-day</td>
<td>1981-1999</td>
<td>11 layers; 0-100 cm</td>
<td>1-9 layers (0-75 cm)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1984-1999</td>
<td>0-50 cm</td>
<td></td>
</tr>
<tr>
<td>PKU</td>
<td>in-situ</td>
<td>% of porosity</td>
<td>station, 10-day</td>
<td>1991-2007</td>
<td>7 layers; 0-100 cm</td>
<td>1-8 layers (0-37 cm)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1992-2006</td>
<td>0-30 cm</td>
<td></td>
</tr>
<tr>
<td>GLEAM</td>
<td>M+RS+RA</td>
<td>m$^3$.m$^{-3}$</td>
<td>0.25°, daily</td>
<td>1980-2014</td>
<td>0-10 cm</td>
<td>1-6 layers (0-9 cm)</td>
</tr>
<tr>
<td>surface</td>
<td></td>
<td></td>
<td></td>
<td>1981-2009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GLEAM</td>
<td>M+RS+RA</td>
<td>m$^3$.m$^{-3}$</td>
<td>0.25°, daily</td>
<td>1980-2014</td>
<td>Mixture of bare soil (1-10 cm), low vegetation (0-100 cm) and</td>
<td>all layers (0-200 cm)</td>
</tr>
<tr>
<td>root-zone</td>
<td></td>
<td></td>
<td></td>
<td>1981-2009</td>
<td>high vegetation (0-250 cm)</td>
<td></td>
</tr>
<tr>
<td>ESA CCI</td>
<td>RS</td>
<td>m$^3$.m$^{-3}$</td>
<td>0.5°, daily</td>
<td>1979-2010</td>
<td>Top layer, depth ≈ 0.5-2 cm</td>
<td>1-4 layers (0-2 cm)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2007-2009</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Median of metrics in specific comparisons. All correlations are significant (p-value < 0.05).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Simulations</th>
<th>Correlation</th>
<th>RMSE (m$^3$.m$^{-3}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISMN</td>
<td>GSWP3</td>
<td>0.51</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>PGF</td>
<td>0.46</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>CRU-NCEP</td>
<td>0.39</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>WFDEI</td>
<td>0.56</td>
<td>0.08</td>
</tr>
<tr>
<td>PKU</td>
<td>GSWP3</td>
<td>0.39</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>PGF</td>
<td>0.31</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>CRU-NCEP</td>
<td>0.32</td>
<td>NA</td>
</tr>
<tr>
<td></td>
<td>WFDEI</td>
<td>0.44</td>
<td>NA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>China</th>
<th>Yangtze</th>
<th>Yellow</th>
<th>China</th>
<th>Yangtze</th>
<th>Yellow</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLEAM</td>
<td>GSWP3</td>
<td>0.57</td>
<td>0.63</td>
<td>0.55</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>surface SM</td>
<td>PGF</td>
<td>0.43</td>
<td>0.53</td>
<td>0.38</td>
<td>0.08</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>CRU-NCEP</td>
<td>0.53</td>
<td>0.61</td>
<td>0.49</td>
<td>0.10</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>WFDEI</td>
<td>0.71</td>
<td>0.80</td>
<td>0.66</td>
<td>0.08</td>
<td>0.08</td>
</tr>
</tbody>
</table>

| GLEAM      | GSWP3 | 0.64    | 0.69   | 0.62  | 0.05    | 0.03   |
| root-zone SM| PGF   | 0.60    | 0.72   | 0.60  | 0.05    | 0.04   |
|            | CRU-NCEP| 0.43  | 0.55   | 0.40  | 0.08    | 0.08   |
|            | WFDEI  | 0.67    | 0.77   | 0.59  | 0.06    | 0.04   |

| ESA CCI    | GSWP3 | 0.43    | 0.38   | 0.51  | 0.06    | 0.06   |
|            | PGF   | 0.27    | 0.29   | 0.27  | 0.06    | 0.07   |
|            | CRU-NCEP| 0.49 | 0.45   | 0.49  | 0.06    | 0.06   |
|            | WFDEI  | 0.54    | 0.49   | 0.60  | 0.06    | 0.06   |
Figures

Figure 1. Map of China. The grey background is cropland fraction. Green rectangulars show four important regions mentioned in this paper.
Figure 2. Mean of annual precipitation (left panel) and simulated total annual SM (right panel) driven by different atmospheric forcing data over 1981-2009. Red and magenta contours are catchment of Yangtze and Yellow River, respectively. Blue lines are river networks.
Figure 3. Pearson correlation coefficients of modeled and measured SM at each gauging station from ISMN (triangles) and PKU (circles). Symbols with dark border indicate significant correlations ($p < 0.05$). The locations of three ISMN stations shown in Figure 4 are marked by black squares in panel (a).
Figure 4. Time series of 10-day SM from ORCHIDEE and ISMN at three stations. The station locations are shown in Fig. 3(a). The mean annual precipitation at Xifeng, Xinxian and Xuzhou (according to GSWP3) are 556, 580 and 847 mm yr$^{-1}$, respectively. Dark dashed lines indicate ISMN SM. Red, green, blue and orange lines indicate simulated SM based on GSWP3, PGF, CRU-NCEP and WFDEI, respectively.
Figure 5. Left panel: Pearson correlation coefficients of the GLEAM surface SM and the corresponding ORCHIDEE SM. Gray indicates non and negative correlation. Right panel: decomposition of the MSD between the daily GLEAM surface SM and the corresponding ORCHIDEE SM (Eq. 1). Cyan, magenta and yellow indicate the fractions of SB, SDSD and LCS respectively.
Figure 6. (a): Number of lag days (NLD) of GLEAM surface SM. (b)-(e): Difference of NLD between GLEAM and ORCHIDEE surface SM. (f)-(h): Autocorrelation coefficient $R_k$ of spatial averaged surface SM as a function of NLD. The dashed line ($y = 1/e$) is the threshold of significant correlation.
Figure 7. Left panel: Correlation coefficients of the ESA CCI SM and the corresponding ORCHIDEE SM. Gray pixels indicate non and negative correlation. Right panel: decomposition of the MSD between the daily ESA CCI SM and the corresponding ORCHIDEE SM (Eq. 1). Cyan, magenta and yellow indicate the fractions of SB, SDSD and LCS respectively.
Figure 8. Evaluation of the forcing datasets for simulating SM dynamics in China, YZRB and YLRB. Top panel: Radar charts of criteria of the four forcing datasets. Center implies bad criteria. Red, green, blue and orange lines indicate GSWP3, PGF, CRU-NCEP and WFDEI, respectively. ‘surf’ and ‘root’ indicates surface and root-zone SM of GLEAM 3.0A. Bottom panel: Composition of MSD from each comparison. x-axis indicates the drivers of specific simulations; top labels indicate the data set used in the specific comparison.
Figure 9. $D$ of simulated SM and meteorological variables. In each grid cell, the monthly time series of specific variable are chosen from the four simulations and the difference is calculated by Eq. 6. The unit of $D$ varies with specific variables. Thus only the spatial patterns can be compared, not the $D$ value.
Figure 10. Matrix of correlation coefficients between the $D$ of meteorological variables and the $D$ of simulated SM. $\theta_{pi}$ indicates SM in layer $i$. $N_{pi}$ indicates the number of precipitation days within a specific precipitation range (Sect. 3.4).