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

Zun Yin[1], Catherine Ottlé[1], Philippe Ciais[1], Matthieu Guimberteau[1,2], Xuhui Wang[1,3,4], Dan Zhu[1], Fabienne Maignan[1], Shushi Peng[3], Shilong Piao[3], Jan Polcher[4], Feng Zhou[3], Hyungjun Kim[5], and CTS members[6]

[1]Laboratoire des Sciences du Climat et de l'Environnement, CNRS-CEA-UVSQ, Gif-sur-Yvette 91191, France
[2]UMR 7619 METIS, Sorbonne Universités, UPMC, CNRS, EPHE, 4 place Jussieu, Paris 75005, France
[3]Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China
[4]Laboratoire de Météorologie Dynamique, UPMC/CNRS, IPSL, Paris 75005, France
[5]Institute of Industrial Science, The University of Tokyo, Tokyo, Japan
[6]Patrice Dumas, Laurent Li, Tao Yang, Xiaoming Feng, Shuai Wang, Xudong Zhou, Pengfei Shi, Hui Yang, Yi Xi and Xuanze Zhang

Correspondence to: vyin@lsce.ipsl.fr

A Figures

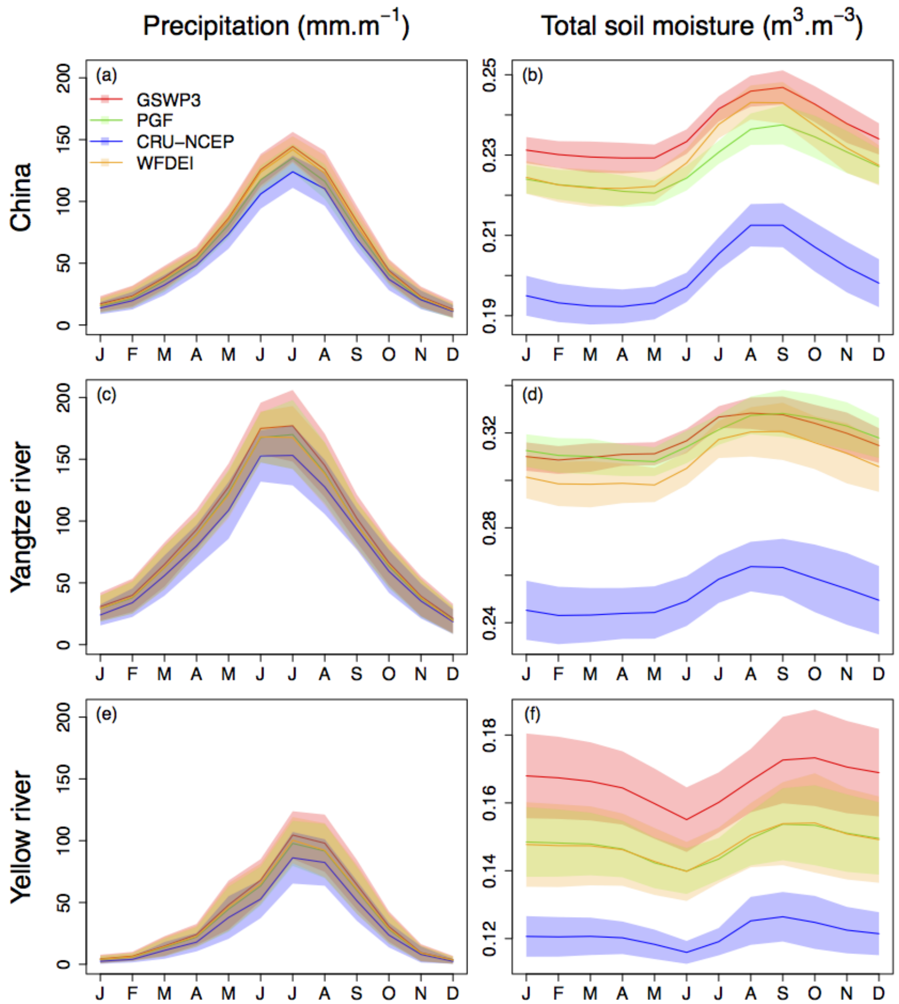

Figure S1. Monthly averaged precipitation (left panel) and total soil moisture (right panel) over 1981–2009. Colors indicate different forcing datasets. Solid lines are the mean and the shadow areas are confidence intervals with one standard deviation.

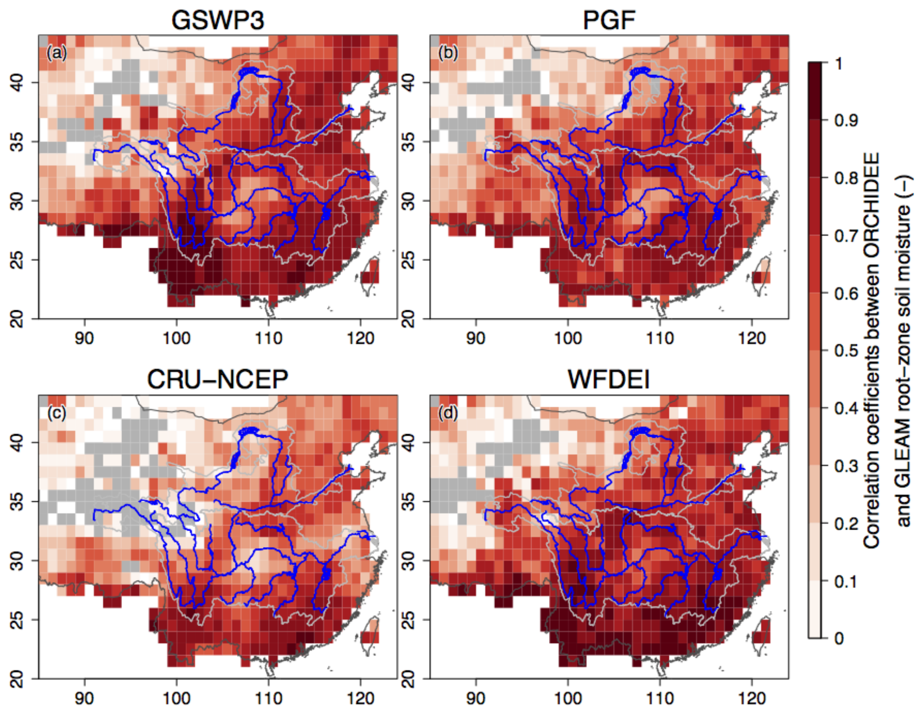

Figure S2. Pearson correlation coefficients between GLEAM root-zone SM and corresponding ORCHIDEE SM. Gray pixels indicate non and negative correlations.

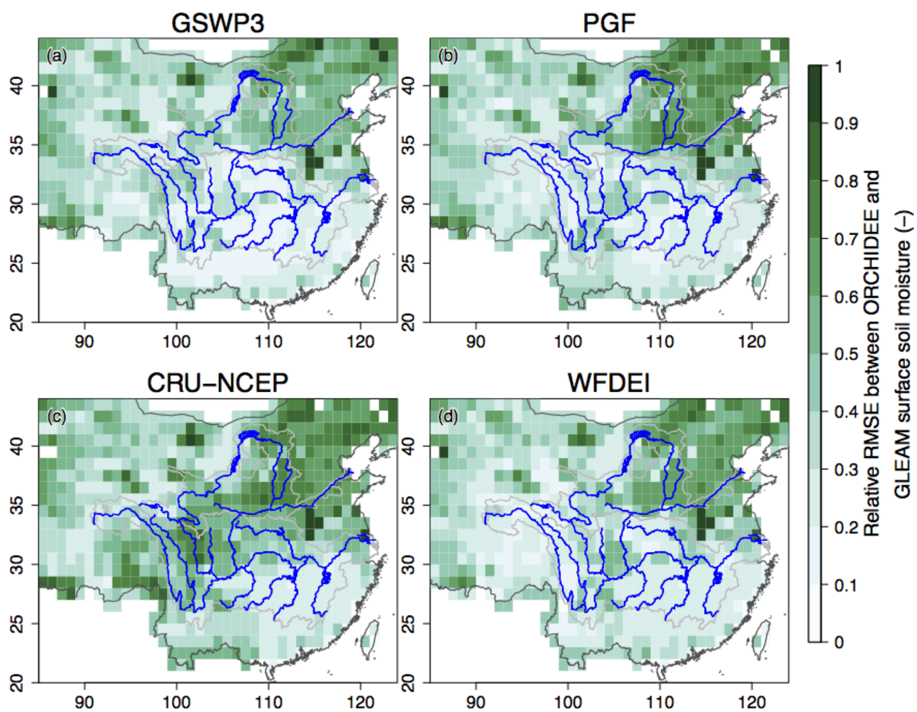

Figure S3. Relative RMSE between the daily GLEAM surface SM and the corresponding ORCHIDEE SM.

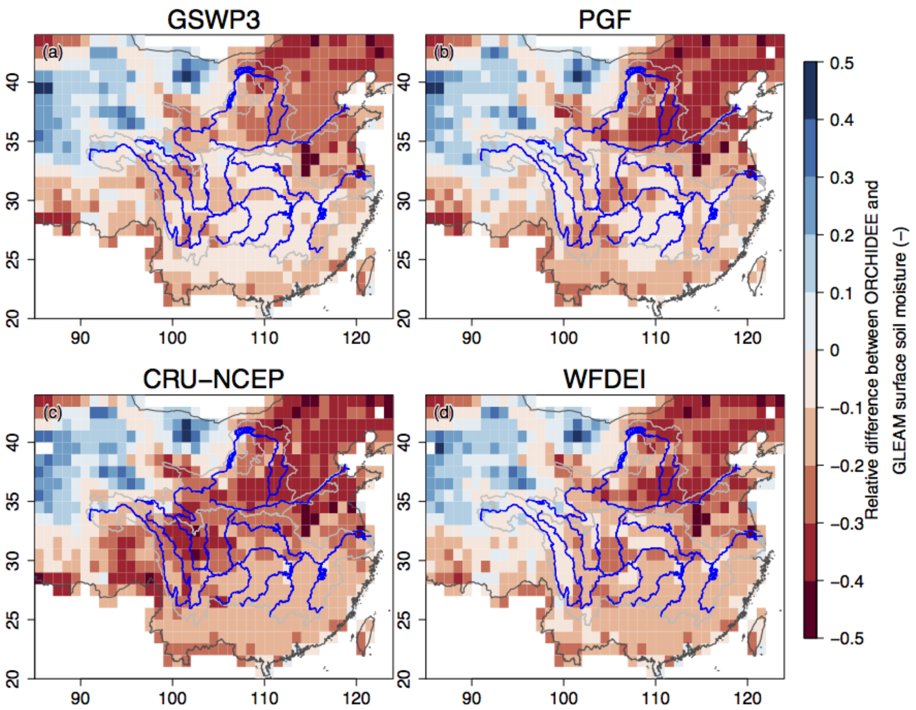

Figure S4. Relative difference ($\Delta_{\text{orc-gleam}}$) between the daily GLEAM surface SM ($\theta_{\text{s,gleam}}$) and the corresponding ORCHIDEE SM ($\theta_{\text{s,orc}}$). The $\Delta_{\text{orc-gleam}}$ is calculated by $(\theta_{\text{s,orc}} - \theta_{\text{s,gleam}})/(\theta_{\text{s,orc}} + \theta_{\text{s,gleam}})$.

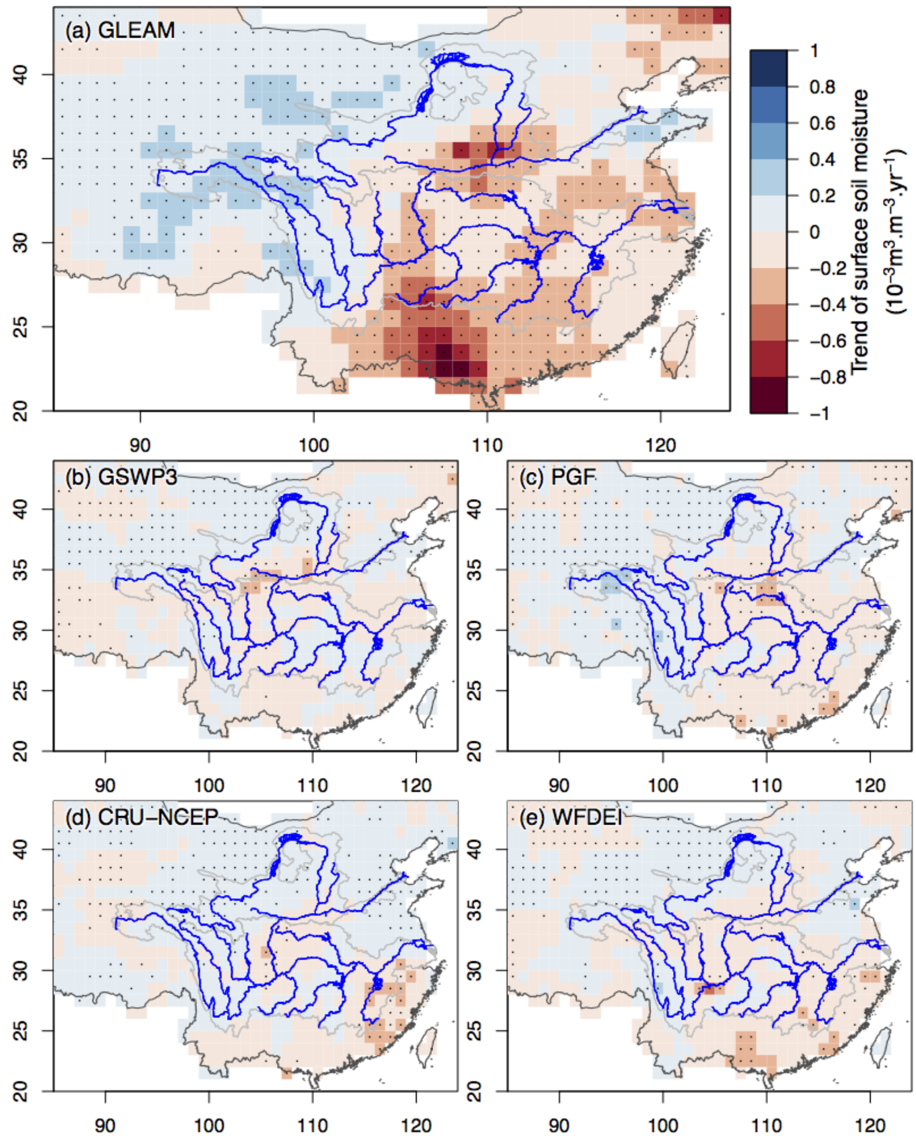

Figure S5. The trend of GLEAM and ORCHIDEE surface SM. Grid cells with dark circles indicate significant trend ($p < 0.05$).

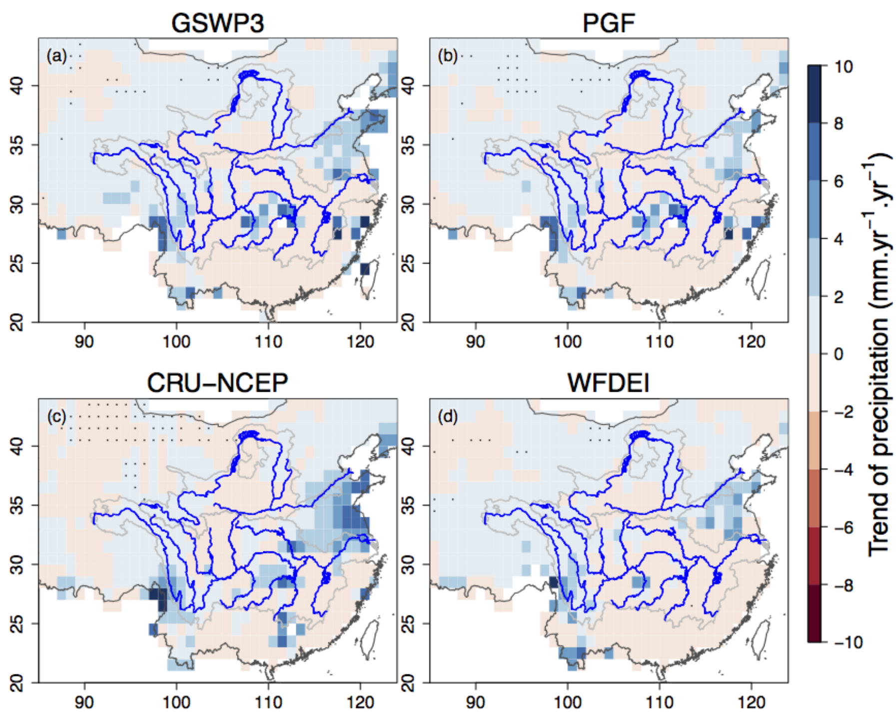

Figure S6. Trend of precipitation form the four forcing datasets. Grid cells with dark circles indicate significant trend ($p < 0.05$).

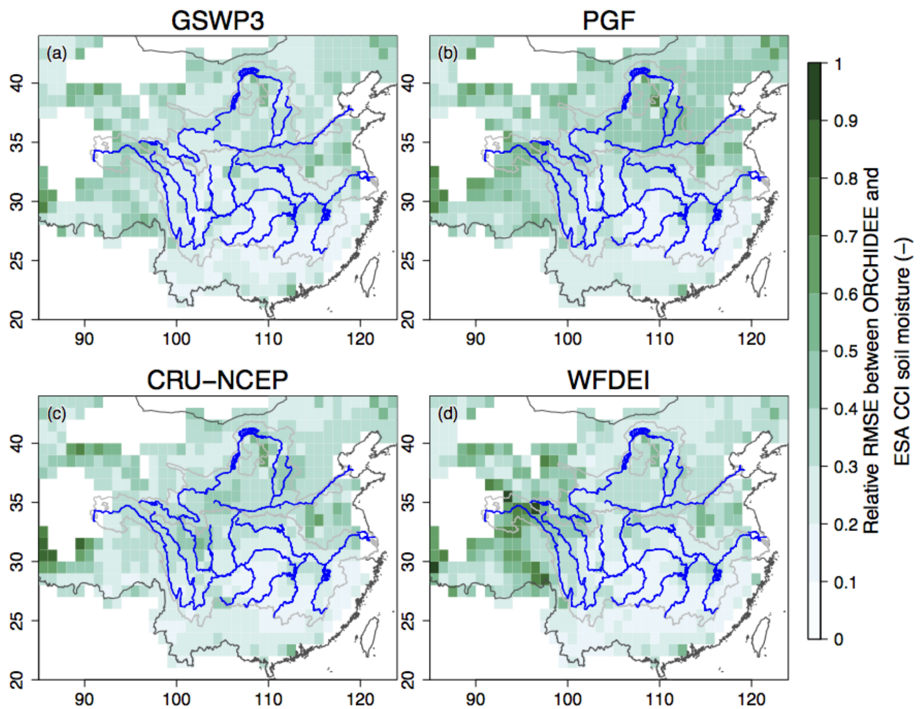

Figure S7. Relative RMSE between the daily ESA CCI SM and the corresponding ORCHIDEE SM.

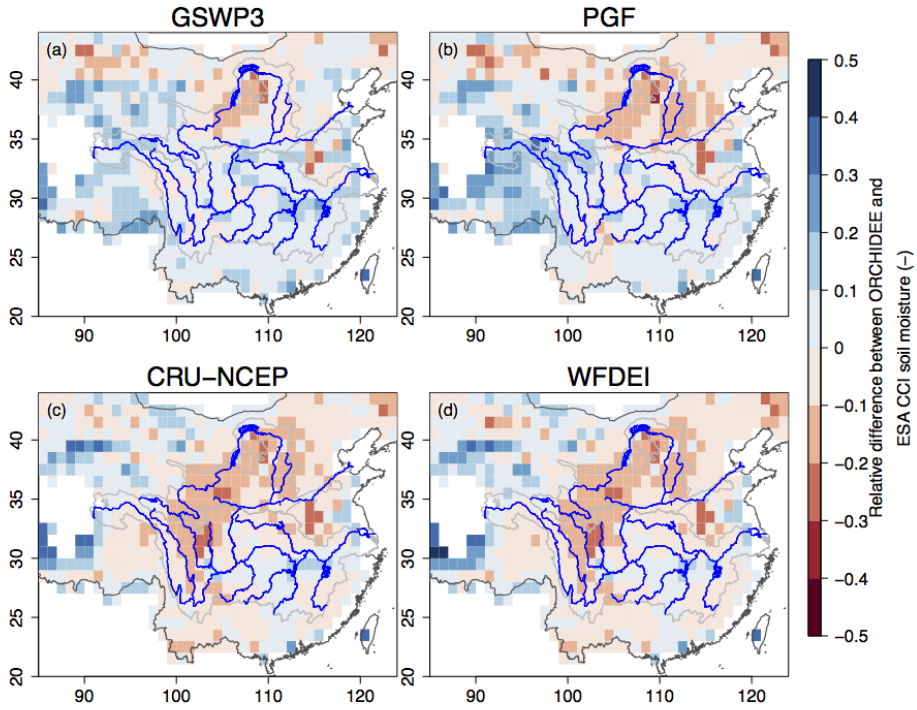

Figure S8. Relative difference ($\Delta_{orc\text{-}cci}$) between the daily ESA CCI SM and the corresponding ORCHIDEE SM.

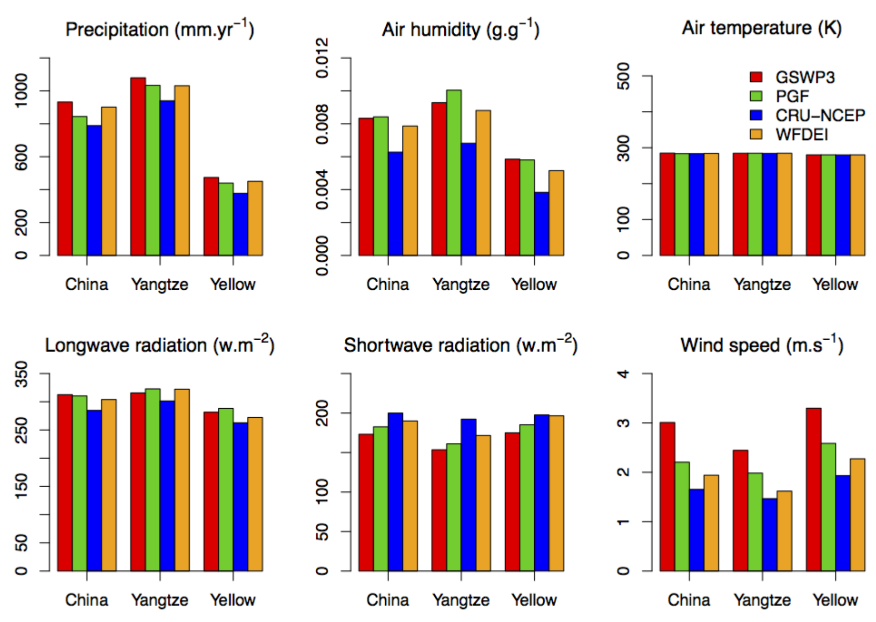

Figure S9. Spatial averaged annual value of meteorological variables from the four forcing datasets in different regions.

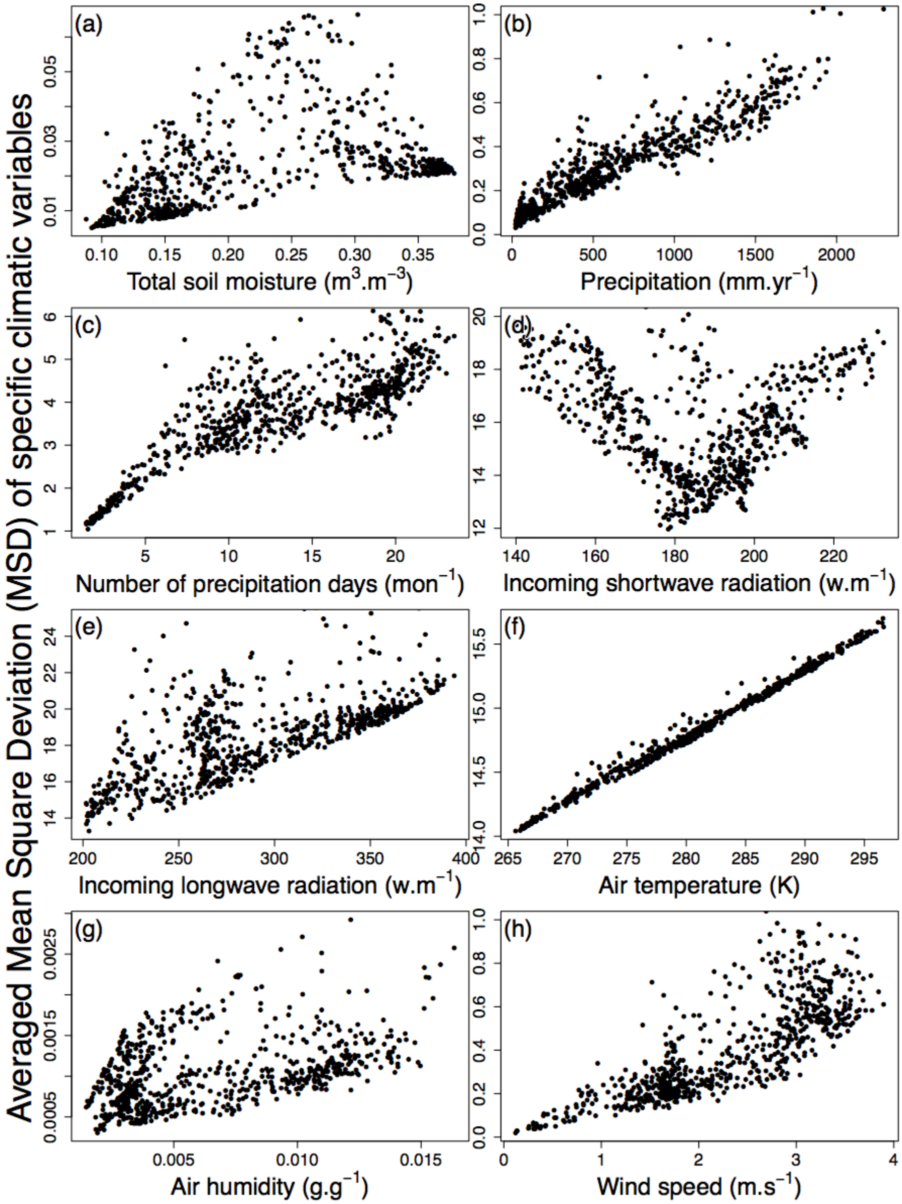

Figure S10. *D* of meteorological variables and simulated SM as a function of the specific variable values.

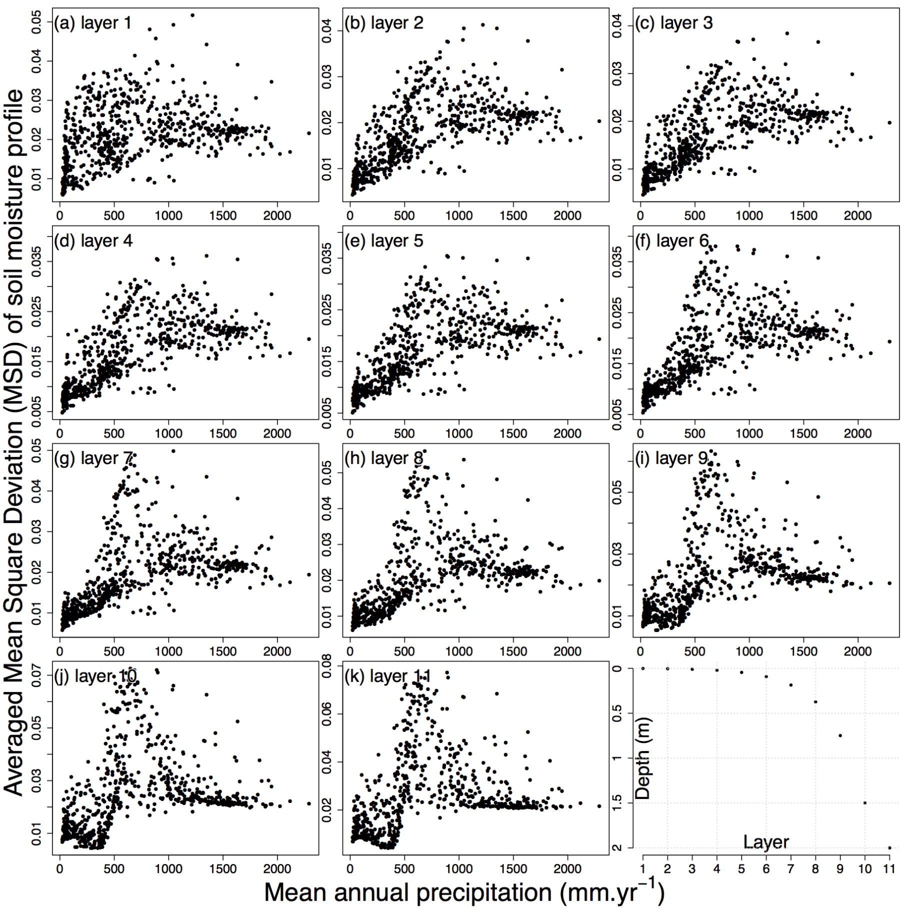

Figure S11. D of SM in different soil layers as a function of mean annual precipitation. The last subplot illustrates the depth of the bottom of each soil layer.

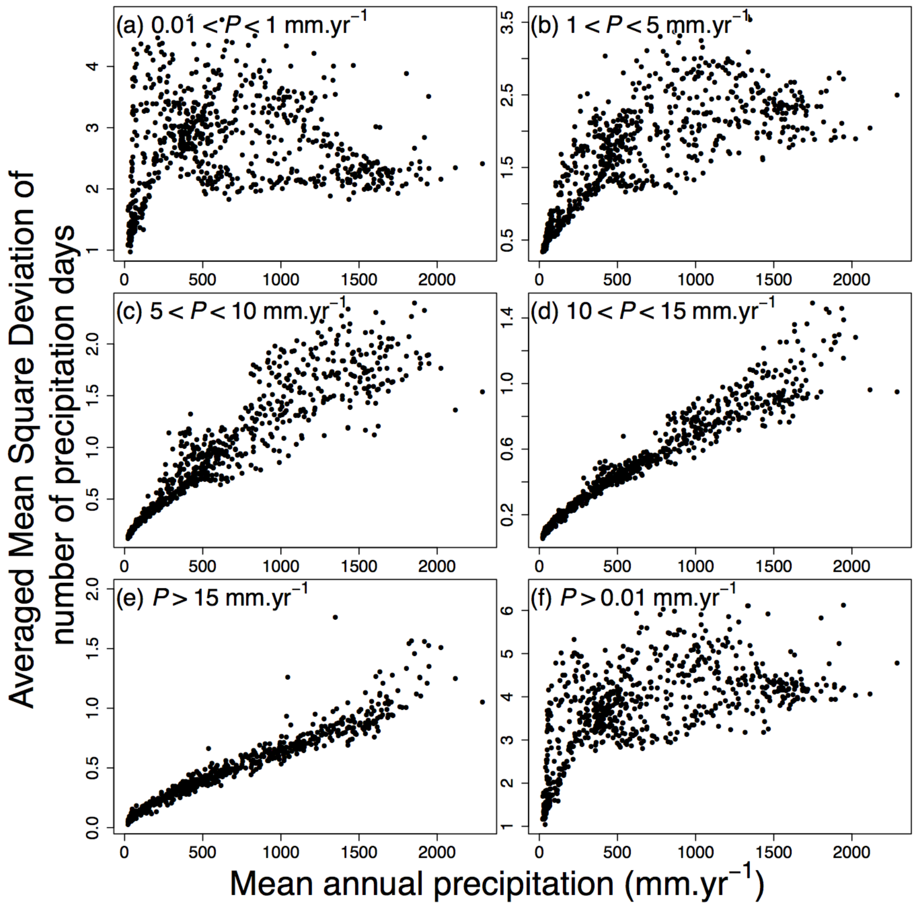

Figure S12. D of N_p at different precipitation bands as a function of mean annual precipitation

B Why relative difference is not suitable?

In this study we want to investigate the difference of which meteorological variable among the four forcing leads to the difference of simulated soil moisture. To address this question, we first calculated the averaged MSD D_x (see Eq. 6) of meteorological variables and SM. Then the correlation coefficients are estimated between them to discover the highest correlation of D between those meteorological variables and simulated SM. However one question is proposed about the methodology: Why not use the relative D for the analysis to remove potential impacts of magnitude of specific variable on its D? Below, we will discuss and compare the absolute and relative D in a simple example to prove that the relative value is not suitable.

We assume that a variable x from atmospheric forcing dominates the simulated variable y through ORCHIDEE. The relation between x and y can be written as a function:

$$y = f(x). \tag{B1}$$

Then for each grid cell m, there are N observations of x^m from N forcing (one observation from one forcing), recorded as $\{x_i^m\}$. The true value of x in grid cell m is $\hat{x}^m$. We assume that the magnitude of observed error correlates with the magnitude of the true value. Then the $\{x_i^m\}$ can be written as:

$$x_i^m = \hat{x}^m + \hat{x}^m \epsilon_i. \tag{B2}$$

Here ϵ_i is Gaussian noise, which obeys the normal distribution with a mean value of 0 and a uniform standard deviation in all grid cells. Then we have:

$$\sum_{i=1}^{N}\sum_{j=1}^{N}(\epsilon_i - \epsilon_j)^2 = 2N^2\sigma^2. \tag{B3}$$

Here σ^2 is the variance of $\{\epsilon_i\}$. We record the value as E^m:

$$\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(\epsilon_i - \epsilon_j)^2}{\binom{N}{2}} = E^m. \tag{B4}$$

Note that E^m varies among the grid cells, which is distributed according to chi-squared distribution. And it is independent to the magnitude of x. The simulated variable y is calculated by Eq. B1. As the ORCHIDEE is a model, with given input, the output is determined by a set of equatons. So the uncertainty is only transferred from the input, as:

$$y_i^m = f(x_i^m). \tag{B5}$$

Now the averaged SMD of $\{x_i^m\}$ can be calculated as:

$$\begin{aligned}
\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(x_i^m - x_j^m\right)^2}{\binom{N}{2}} &= A\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\hat{x}^m + \hat{x}^m\epsilon_i - \left(\hat{x}^m + \hat{x}^m\epsilon_j\right)\right]^2 \\
&= A\left(\hat{x}^m\right)^2\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\epsilon_i - \epsilon_j\right)^2 \\
&= \left(\hat{x}^m\right)^2 E^m
\end{aligned} \tag{B6}$$

Here A is used to temporally record constants. As the same as Eq. B6, the averaged MSD of $\{y_i^m\}$ can be calculated as:

$$\begin{aligned}
\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(y_i^m - y_j^m\right)^2}{\binom{N}{2}} &= A\sum_{i=1}^{N}\sum_{j=1}^{N}\left[f\left(\hat{x}^m + \hat{x}^m\epsilon_i\right) - f\left(\hat{x}^m + \hat{x}^m\epsilon_j\right)\right]^2 \\
&\approx A\sum_{i=1}^{N}\sum_{j=1}^{N}\left[\left(f\left(\hat{x}^m\right) + f'\left(\hat{x}^m\right)\hat{x}^m\epsilon_i\right) - \left(f\left(\hat{x}^m\right) + f'\left(\hat{x}^m\right)\hat{x}^m\epsilon_j\right)\right]^2 \\
&= A\sum_{i=1}^{N}\sum_{j=1}^{N}\left[f'\left(\hat{x}^m\right)\hat{x}^m\left(\epsilon_i - \epsilon_j\right)\right]^2 \\
&= \left[f'\left(\hat{x}^m\right)\hat{x}^m\right]^2 E^m
\end{aligned} \tag{B7}$$

It is clear that the D of output $\{y_i^m\}$ depends on the magnitude of $\{x_i^m\}$, not on the magnitude of output. In other words, the Ds of input and output are correlated due to such uncertainty transport through a model. But the relative Ds, D over the value of specific variable, are probably not correlated as they removed the magnitude information of input and brought extra uncertainty of output.