Improving the simulation of terrestrial water storage anomalies over China using a Bayesian model averaging ensemble approach

LIU Jian-Guo\textsuperscript{ab}, JIA Bing-Hao\textsuperscript{b}, XIE Zheng-Hui\textsuperscript{b} and SHI Chun-Xiang\textsuperscript{c}

\textsuperscript{a}School of Mathematics and Computational Science, and Key Laboratory of Intelligent Control Technology for Wuling-Mountain Ecological Agriculture in Hunan Province, Huaibei University, Huaibei, China; \textsuperscript{b}State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics (LASG), Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China; \textsuperscript{c}National Meteorological Information Center, China Meteorological Administration, Beijing, China

ABSTRACT

The ability to estimate terrestrial water storage (TWS) is essential for monitoring hydrological extremes (e.g., droughts and floods) and predicting future changes in the hydrological cycle. However, inadequacies in model physics and parameters, as well as uncertainties in meteorological forcing data, commonly limit the ability of land surface models (LSMs) to accurately simulate TWS. In this study, the authors show how simulations of TWS anomalies (TWSAs) from multiple meteorological forcings and multiple LSMS can be combined in a Bayesian model averaging (BMA) ensemble approach to improve monitoring and predictions. Simulations using three forcing datasets and two LSMS were conducted over mainland China for the period 1979–2008. All the simulations showed good temporal correlations with satellite observations from the Gravity Recovery and Climate Experiment during 2004–08. The correlation coefficient ranged between 0.5 and 0.8 in the humid regions (e.g., the Yangtze river basin, Huaibei basin, and Zhuijiang basin), but was much lower in the arid regions (e.g., the Heihe basin and Tarim river basin). The BMA ensemble approach performed better than all individual member simulations. It captured the spatial distribution and temporal variations of TWSAs over mainland China and the eight major river basins very well; plus, it showed the highest R value (> 0.5) over most basins and the lowest root-mean-square error value (< 40 mm) in all basins of China. The good performance of the BMA ensemble approach shows that it is a promising way to reproduce long-term, high-resolution spatial and temporal TWSA data.

1. Introduction

Terrestrial water storage (TWS), which includes soil moisture (SM), groundwater storage (GWS), surface water (SW), vegetation water (VW), snow and ice, plays a fundamental role in the global hydrological cycle (Philippon and Fontaine 2002; Grippa et al. 2011). TWS anomalies (TWSAs) can affect the hydrological cycle through SM feedbacks to the atmosphere. TWS acts as a ‘memory’ component of the terrestrial hydrological cycle within the climate system through land–atmosphere interactions due to its unique dynamics (Koster and Suarez 2001). There is a strong correspondence between TWSAs and extreme hydrological events (e.g., droughts and floods). It is thus very important to obtain long-term, accurate, high-resolution spatial and temporal TWSA information to monitor and predict extreme hydrological events, as well as to evaluate the potential impact of these variations on climate change (Dirmeyer, Schlosser, and Brubaker 2009; Chen et al. 2010; Koster et al. 2010).

Three methods are traditionally used to obtain TWSA estimates: satellite observations of gravity anomalies, such as the Gravity Recovery and Climate Experiment (GRACE) satellite data (Syed et al. 2008; Chen et al. 2010); the water balance method (Zeng et al. 2007); and land surface models (LSMs) (Grippa et al. 2011; Sun, Xie, and Tian 2015). Each method presents advantages and
disadvantages. At present, mechanistic LSMS are widely used to estimate land hydrological variables such as SM, DW, and evapotranspiration at global and continental scales (Wang and Zeng 2011; Wang and Dickinson 2012); however, LSM simulations of land hydrological variables are still very inaccurate. The main sources of error in offline LSM simulations of land hydrological variables are the uncertainties in both meteorological forcing data and land surface parameterizations (Dirmeyer, Dolman, and Sato 1999; Wang and Zeng 2011; Wang, Lettenmaier, and Sheffield 2011). Ensemble simulations combining multiple LSMS, or multiple forcings, have been found to perform better than a single LSM or forcing dataset in most cases (Guo et al. 2007; Liu et al. 2016).

However, whilst ensemble simulations combining multiple LSMS or multiple forcings using the simple arithmetical average ensemble method have been shown to outperform most individual member simulations over most subregions, they were still found to be inferior to the best individual ensemble member in most cases (Guo et al. 2007; Liu and Xie 2013).

Bayesian model averaging (BMA) was proposed by Raftery et al. (2005) as a statistical post-processing procedure that infers consensus predictions by weighting individual member predictions based on their probabilistic likelihood measures, with the better performing predictions receiving higher weights in the ensemble predictions (Duan et al. 2007; Li, Zeng, and Li 2009). Many studies have demonstrated that this approach is superior to the simple arithmetical averaging method, and provides a quantitative description of total predictive uncertainty through the probability density function (PDF), especially in numerical weather prediction (Raftery et al. 2005; Duan et al. 2007; Liu and Xie 2013; Li and Lin 2015; Jia and Xie 2016). However, few studies have focused on applying the BMA ensemble approach to multiple LSMS and multiple forcing ensemble simulations. An approach considering the uncertainties from the forcings and LSM simultaneously has the potential to improve TWSA simulations. In the present work, four simulations using three meteorological forcings to drive two LSMS over mainland China were merged using the BMA ensemble approach, and a comparison was conducted between the four individual simulations, the BMA ensemble simulations, and the GRACE satellite observed TWSA data.

2. Models, data and methods

2.1. Meteorological forcings

In this study, we used three meteorological forcing datasets developed by different institutions to drive LSMS over mainland China: (1) Qian (Qian et al. 2006; Tian et al. 2010); (2) Princeton (Sheffield, Goteti, and Wood 2006); (3) Institute of Tibetan Plateau Research, Chinese Academy of Sciences (ITPCAS) (He 2010). The three meteorological forcing datasets are the same as employed in Liu et al. (2016). Table 1 in Liu et al. (2016) summarizes the primary features and the differences between these forcing datasets.

2.2. LSMS

The two LSMS used in this study were version 3.5 of the Community Land Model (CLM3.5) (Oleson et al. 2007), and version 4.5 (CLM4.5) (Oleson et al. 2013), released by the National Center for Atmospheric Research. In these two LSMS, TWS is the sum of GWS, SM, snow water equivalent (SWE), VW, and SW. These hydrological variables are the main variables in land hydrological processes. The TWSA simulated by the two LSMS assumes that the contribution of VW and SW are

Table 1. Comparison of the major features of LSM structures and hydrological schemes.

| Model   | Major hydrology scheme                                                                 | Reference                                      |
|---------|----------------------------------------------------------------------------------------|------------------------------------------------|
| CLM3.5  | (1) The total soil column of 0–3.43 m is divided into 10 layers with variable thickness ranging from 0.0175 to 1.1370 m | Oleson et al. (2007); Liu et al. (2005, 2007); Lawrence and Chase (2007) |
|         | (2) Scaling of canopy interception                                                      |                                                 |
|         | (3) A combination of the TOPMODEL and BATS runoff scheme                                |                                                 |
|         | (4) A snow submodel with up to five layers depending on the total snow depth           |                                                 |
|         | (5) A simple groundwater model for determining water table depth                       |                                                 |
| CLM4.5  | (1) The total soil column 0–42.1032 m is divided into 15 layers with variable thickness ranging from 0.0175 to 13.8512 m | Lawrence et al. (2012); Swenson and Lawrence (2012); Oleson et al. (2013) |
|         | (2) New treatments of soil column-groundwater interactions and soil evaporation         |                                                 |
|         | (3) Revised hydraulic properties of frozen soils, increasing the consistency between soil water state and water table position and allowing for a perched water table above icy permafrost ground |                                                 |
|         | (4) Revised snow model                                                                 |                                                 |
|         | (5) Introduction of a surface water store                                              |                                                 |
|         | (6) Refinement of global PFT, wetland, and lake distributions, and more realistic optical properties for grasslands and croplands |                                                 |
|         | (7) Addition of an urban canyon model                                                  |                                                 |
negligible. In this study, simulated TWS values in each individual member model were converted to TWSA by subtracting the 1979–2008 modeled mean TWS.

Table 1 summarizes the primary features and the differences between the two LSMs, as well as the sources of the vegetation and soil parameters.

2.3. GRACE data

Due to the lack of direct TWSA observations, we used GRACE data to compare and evaluate our five TWSA simulations (i.e., four individual member simulations and their BMA ensemble simulation). Since the GRACE satellite launched in 2002, many studies have used GRACE data for various hydrological applications (e.g., Chen et al. 2010; Grippa et al. 2011). The GRACE satellite mission provides observations of land gravity changes, from which TWS changes can be derived. Different methods can be used to retrieve the TWS from satellite gravity observations, and these different methods result in different GRACE TWSA datasets (Rowlands et al. 2005; Chambers 2006; Bruinsma et al. 2010). In our study, the GRACE-derived TWSA data were provided by the Tellus product processed by NASA’s Jet Propulsion Laboratory, which were smoothed with a 300 km radius and converted to a grid format at a monthly temporal and 1° × 1° spatial resolution (Wahr et al. 2004; Swenson and Wahr 2006; Sun, Xie, and Tian 2015). The five-year complete dataset for 2004–08 was used for this study.

2.4. Experimental setup

The four individual ensemble member TWSA simulations were derived from two different LSMs (described in Table 1) driven by three different forcing datasets, which we coded as: (1) Qian_CLM3.5; (2) Prin_CLM3.5; (3) ITP_CLM3.5; and (4) ITP_CLM4.5. These simulations were all run at a resolution of 1° × 1°. The four sets of simulated TWSA data were then merged using the BMA ensemble method, referred to simply as ‘BMA’. Table 2 gives more specific information on the experimental design.

2.5. Ensemble approach

In this study, an advanced BMA ensemble approach was used. The BMA approach employed in this study is the same as employed in Liu and Xie (2013). The description of the method parallels that of Liu and Xie (2013), as follows in the next two paragraphs.

The BMA approach is a weighted average of PDFs centered on the bias-corrected predictions from a set of individual ensemble members:

\[ p(y|(f_1, \cdots, f_K, y^T)) = \sum_{k=1}^{K} w_k p_k(y|(f_k, y^T)), \]

where \( y \) is the predictive variable; \( f_k, k = 1, 2, \cdots, K \) is the \( k \)th ensemble member forecasts; \( K \) is the number of ensemble members being combined; \( y^T \) are the training data; and \( p_k(y|(f_k, y^T)) \) is the conditional PDF of \( y \) based on ensemble member \( f_k \), given that \( f_k \) is the best forecast in the ensemble. The weighting \( w_k \) is the posterior probability of the forecast, such that it is non-negative and \( \sum_{k=1}^{K} w_k = 1 \), and represents the contribution of ensemble member \( k \) to the predictive skill of the ensemble. In the BMA we developed and used, we assumed that the conditional PDF \( p_k(y|(f_k, y^T)) \) from each ensemble member at the specific time and location was approximated as a gamma distribution:

\[ p_k(y|(f_k, y^T)) = \frac{1}{\beta_k^a \Gamma(a_k)} y^{a_k - 1} \exp(-y/\beta_k), \]

where \( \Gamma(a_k) \) is the gamma function. The shape parameter \( a_k \) and scale parameter \( \beta_k \) of the gamma distribution are given by:

\[ \mu_k = a_k \beta_k = b_{0k} + b_{1k} f_k, \]
\[ \sigma_k^2 = a_k \beta_k^2 = c_0 + c_1 f_k, \]

where \( \mu_k \) and \( \sigma_k^2 \) are the mean and variance of the gamma distribution, and \( b_{0k}, b_{1k}, c_0, c_1 \) are the parameters to be estimated. The mean of the BMA posterior PDF is the deterministic forecast. In this study, we

| Table 2. Specific information regarding the experimental design. |
|-------------------------------------------------------------|
| Name of experiment | Forcing | LSM |
|---------------------|---------|-----|
| Qian_CLM3.5         | Qian: Forcing developed by Qian et al. (2006) and Tian et al. (2010) | Community Land Model, version 3.5 (CLM3.5) |
| Prin_CLM3.5         | Princeton: Forcing developed by Sheffield, Goteti, and Wood (2006) from Princeton University | CLM3.5 |
| ITP_CLM3.5          | ITPCAS: Forcing developed by He (2010) from the Institute of Tibetan Plateau Research, Chinese Academy of Sciences (ITPCAS) | CLM3.5 |
| ITP_CLM4.5          | ITPCAS | Community Land Model, version 4.5 (CLM4.5) |
| BMA                 | BMA merged TWSA data from Qian_CLM3.5, Prin_CLM3.5, ITP_CLM3.5, and ITP_CLM4.5 | |
considered only the deterministic forecast of the BMA method for the TWSA.

The BMA weights $\omega_k$ and parameters $b_{0k}, b_{1k}, c_0, c_1$ in this case were estimated from a training dataset using the linear regression method and the maximum likelihood technique, for which the values were obtained iteratively using the modified Markov chain Monte Carlo algorithm. In the present study, 2004–06 was chosen as the training period, and 2007–08 as the evaluation period.

3. Results

3.1. Spatial distribution of TWSAs

To examine the performance of the BMA ensemble simulation of TWSAs, we first compared the four simulations and their BMA ensemble simulation of the mean TWSA with GRACE data from 2004–08. Figure 1 shows the spatial distribution of the five-year (2004–08) averaged TWSA in China derived from Qian_CLM3.5, Prin_CLM3.5, ITP_CLM3.5, ITP_CLM4.5, BMA, and GRACE data. The GRACE data exhibited strong regional variations and there were several obvious negative anomaly centers, such as the Beijing–Tianjin–Hebei region, the northern region of Northeast China, and the northwestern region of Xinjiang Province; and positive anomaly centers, such as the southwestern region of China and the mid-northern Tibetan Plateau (Figure 1(f)). Among all individual member simulations, only ITP_CLM3.5 captured these negative and positive anomaly centers, although the magnitude differed from the GRACE data. Other individual member simulations did not follow the pattern observed in the GRACE data (Figure 1(a–d)). The BMA ensemble simulation captured the TWSA spatial distribution pattern in China very well, and with similar magnitude, compared with the GRACE data (Figure 1(e)).

3.2. TWSA temporal variation

To evaluate the performance of the BMA ensemble simulated TWSA quantitatively, we compared the individual model outputs, BMA, and GRACE data time series averaged over the eight major basins of mainland China for the period 2004–08. Figure 2 shows the comparisons between ITP_CLM3.5, BMA, and GRACE data. Due to the large difference in magnitude between Qian_CLM3.5, Prin_CLM3.5, ITP_CLM4.5 and the GRACE data, Qian_CLM3.5, Prin_CLM3.5, ITP_CLM4.5 and ITP_CLM4.5 are not shown in Figure 2.

Figure 1. Spatial distribution of terrestrial water storage anomalies (units: mm) during 2004–08, from (a) Qian_CLM3.5, (b) Prin_CLM3.5, (c) ITP_CLM3.5, (d) ITP_CLM4.5, (e) BMA, and (f) GRACE.
ITP_CLM3.5 and BMA captured the seasonal cycle and temporal evolution of the GRACE data very well over most basins, but ITP_CLM3.5 showed a higher bias than the BMA ensemble simulation. Qian_CLM3.5, Prin_CLM3.5, and ITP_CLM4.5 also captured the seasonal cycle of the GRACE data well over most basins, but showed a higher bias compared to ITP_CLM3.5. It should be noted that LSM simulations usually reproduce anomalies and seasonal variations but fail to simulate the mean SM (Qian et al. 2006; Liu and Xie 2013). By contrast, the BMA ensemble modeling did simulate the mean TWSA value.

3.3. Statistical comparisons between simulations and GRACE data

To further evaluate the performance of the BMA ensemble approach in simulating TWSAs, we compared the correlation coefficient ($R$), root-mean-square error (RMSE), and normalized standard deviation (SDV) of the five simulations. $R$ and SDV were plotted on two-dimensional Taylor diagrams (Taylor 2001). The statistical post-processing component of the BMA approach requires training data to calibrate the BMA model parameters. We chose 2004–06 as the training period and 2007–08 as the evaluation period.

Figure 3 shows the statistical scores ($R$ and RMSE) for Prin_CLM3.5, ITP_CLM3.5, ITP_CLM4.5, and BMA over the validation period (2007–08) in the eight major basins of China (due to the differences in magnitude, Qian_CLM3.5 is not shown in Figure 2). In the humid areas (e.g., the

![Figure 2](image_url1)  
*Figure 2.* Time series for terrestrial water storage anomalies (units: mm), obtained from ITP_CLM3.5, BMA, and the GRACE data, for the eight major river basins of China.

![Figure 3](image_url2)  
*Figure 3.* Statistical scores for Prin_CLM3.5 (P35), ITP_CLM3.5 (T35), ITP_CLM4.5 (T45), and BMA, in the validation period 2007–08, in the eight major basins of China: (a) correlation coefficient; (b) RMSE.
Yangtze river basin, Huaihe basin, and Zhujiang basin), all the simulations showed good temporal correlation ($R$), ranging from 0.5 to 0.8. However, the models performed worse in the arid areas (e.g., the Heihe basin and Tarim river basin) than in the humid areas. This may be explained by the impact of human activities, such as the extraction of groundwater for domestic, industrial, and agricultural water consumption. In the individual member experiments, no single model performed optimally in all the basins. The BMA ensemble simulation showed the highest $R$ value (> 0.5) over most basins of China and the smallest RMSE value (< 40 mm) over all the basins of China, and performed better than all individual member simulations.

Figure 4 shows a Taylor diagram comparing the five experiments with the observed GRACE data over the eight major basins of China. ITP_CLM3.5 generally performed best among the four individual member simulations, and the BMA ensemble simulation performed best among all five experiments in most cases. The BMA ensemble approach improved the quality of the simulated TWSA significantly, not only by improving the simulation of the spatial (Figure 1) and temporal (Figure 2) variations, but also by reducing the mean bias (Figure 3).

4. Discussion

Uncertainties in terrestrial water cycle simulations by offline LSMs result from uncertainties in LSM parameterizations, meteorological forcings, and land surface parameter information. We need to reduce these uncertainties in order to improve terrestrial water cycle modeling. The BMA ensemble approach shows promise as a way of reproducing long-term, high-resolution, spatial and temporal TWSA data, and could be useful in supplementing the current lack of long-term TWSA data.

Some limitations of our study should be noted. Firstly, the BMA method depends on the accuracy of observations to provide an adequate training dataset for calibrating the model parameters. We assumed that the GRACE data were true values. In fact, different retrieval methods applied to the land gravity changes detected by GRACE result in different GRACE TWSA data. These uncertainties were not quantified in this study. Secondly, the TWSA is the sum of changes in GWS, SM, SWE, VW, and SW. The LSMs in this study neglected the contribution of VW and SW in TWSA estimates. Thirdly, the approach considers uncertainties in LSM parameterization and meteorological forcing, but does not consider uncertainties in land surface parameter information.

In this study, we considered only the deterministic forecast of the BMA method for the TWSA. The BMA approach can produce calibrated and sharply predictive PDFs from the ensembles of dynamic models and provide probability forecasting, as pointed out by Liu and Xie (2014) and Li and Lin (2015). In future work, we intend to extend the BMA TWSA simulation to probability prediction, and investigate the warning schemes of extreme events such as floods and droughts, based on the BMA TWSA prediction.

5. Summary and conclusion

In this study, we conducted simulations using two LSMs and three sets of meteorological forcing data over mainland China during the period 1979–2008. The four TWSA outputs from the simulations were then merged using a BMA ensemble approach. We compared these five simulated TWSA datasets with GRACE TWSA satellite data over mainland China and the eight major river basins of China during 2004–08. We evaluated the extent to which the quality of TWSA modeling was improved by using the BMA ensemble approach. Our main conclusions are as follows:

Of the four individual member simulations of the TWSA using the three individual forcings and two LSMs, ITP_CLM3.5 best captured the spatial pattern and temporal variation of the TWSA over most basins of mainland China, and performed better than the other three individual member simulations, but it still generated mean bias when compared with the GRACE data. In general, ITP_CLM3.5 performed best and Qian_CLM3.5 performed worst in most basins. This
result is associated with the quality of the meteorological forcing and land surface parameterizations.

The BMA ensemble approach combining the four individual member simulations significantly improved the ability to accurately simulate the TWSA. Not only does the approach reproduce the spatial and temporal variation of the TWSA, but it can also model the mean value of the TWSA. This approach is a promising way to accurately model the mean value and variations of the TWSA in China.

Acknowledgments

We thank Prof. Kun YANG for providing the ITPCAS meteorological forcing data.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This research was supported by the National Natural Science Foundation of China (Grant Nos. 41405083 and 91437220), the Natural Science Foundation of Hunan Province, China (Grant No. 2015JJ3098), the Key Research Program of Frontier Sciences, CAS (QYZDY-SSW-DQC012), and the Fund Project for The Education Department of Hunan Province (Grant No. 16A234).

References

Bruinsma, S., J. M. Lemoine, R. Biancale, and N. Valès. 2010. “CNES/GRGS 10-Day Gravity Field Models (Release 2) and Their Evaluation.” Advances in Space Research 45: 587–601. doi:10.1016/j.asr.2009.10.012.

Chambers, D. P. 2006. “Evaluation of New GRACE Time-Variable Gravity Data over the Ocean.” Geophys Research Letters 33: L17603. doi:10.1029/2006GL027296.

Chen, J. L., C. R. Wilson, B. D. Tapley, L. Longuevergne, Z. L. Yang, and B. R. Scanlon. 2010. “Recent La Plata Basin Drought Conditions Observed by Satellite Gravimetry.” Journal Geophys Researcher 115: D22108. doi:10.1029/2010JD014689.

Dirmeyer, P. A., A. J. Dolman, and N. Sato. 1999. “The Pilot Phase of the Global Soil Wetness Project.” Bulletin of the American Meteorological Society 80 (S): 851–878. doi:10.1175/1520-0477(1999)080<0851:TPPOGT>2.0.CO;2.

Dirmeyer, P. A., C. A. Schlosser, and K. L. Brubaker. 2009. “Precipitation, Recycling, and Land Memory: An Integrated Analysis.” Journal of Hydrometeorology 10: 278–288. doi:10.1175/2008JHM1016.1.

Duan, Q. Y., N. K. Ajami, X. G. Gao, and S. Sorooshian. 2007. “Multi-Model Ensemble Hydrologic Prediction Using Bayesian Model Averaging.” Advances in Water Resources 30: 1371–1386. doi:10.1016/j.advwatres.2006.11.014.

Grippa, M., L. Kergoat, F. Frappart, Q. Araud, A. Boone, P. De Rosnay, J. M. Lemoine, et al. 2011. “Land Water Storage Variability over West Africa Estimated by Gravity Recovery and ClimateExperiment (GRACE) and Land Surface Models.” Water Resources Researcher 47: W05549. doi:10.1029/2009WR008856.

Guo, Z. C., P. A. Dirmeyer, X. Gao, and M. Zhao. 2007. “Improving the Quality of Simulated Soil Moisture with a Multi-Model Ensemble Approach.” Quarterly Journal of the Royal Meteorological Society 133: 731–747. doi:10.1002/qj.1477-870X.

He, J. 2010. “Development of Surface Meteorological Dataset of China with High Temporal and Spatial Resolution.” M.S. thesis, Inst. of Tibetan Plateau Res., Chin. Acad. of Sci., Beijing, China.

Jia, B., and Z. Xie. 2016. “Improving Microwave Brightness Temperature Predictions Based on Bayesian Model Averaging Ensemble Approach.” Applied Mathematics and Mechanics 37 (11): 1501–1516. doi:10.1007/s10483-016-2103-6.

Koster, R. D., S. P. P. Mahanama, B. Livneh, D. P. Lettenmaier, and R. H. Reichle. 2010. “Skill in Streamflow Forecasts Derived from Large-Scale Estimates of Soil Moisture and Snow.” Nature Geoscience 3: 613–616. doi:10.1038/ngeo944.

Koster, R. D., and M. J. Suarez. 2001. “Soil Moisture Memory in Climate Models.” Journal of Hydrometeorology 2: 558–570. doi:10.1175/1525-7541(2001)002<0558:SMMICM>2.0.CO;2.

Lawrence, P. J., and T. N. Chase. 2007. “Representing a New MODIS Consistent Land Surface in the Community Land Model (CLM 3.0).” Journal of Geophysical Research 112. doi:10.1029/2006JG000168.

Lawrence, P. J., J. J. Feddema, G. B. Bonan, G. A. Meehl, B. C. O’Neill, K. W. Oleson, S. Levis, et al. 2012. “Simulating the Biogeoclimatic and Biogeophysical Impacts of Transient Land Cover Change and Wood Harvest in the Community Climate System Model (CCSM4) from 1850 to 2100.” Journal of Climate 25: 3071–3095. doi:10.1175/JCLI-D-11-00256.1.

Li, F., and D. L. Lin. 2015. “Improving Multi-Model Ensemble Probabilistic Prediction of Yangtze River Valley Summer Rainfall.” Advances in Atmospheric Sciences 32: 497–504. doi:10.1007/s00376-014-4073-8.

Li, F., Q. C. Zeng, and C. F. Li. 2009. “A Bayesian Scheme for Probabilistic Multi-Model Ensemble Prediction of Summer Rainfall over the Yangtze River Valley.” Atmospheric and Oceanic Science Letters 2: 314–319. doi:10.1080/16742834.2009.11446815.

Liu, J. G., B. H. Jia, Z. H. Xie, and C. X. Shi. 2014. “Advancement of Surface Meteorological Dataset of China with High Temporal and Spatial Resolution.” M.S. thesis, Inst. of Tibetan Plateau Res., Chin. Acad. of Sci., Beijing, China.

Li, F., Z. D. Lin. 2015. “Improving Multi-Model Ensemble Probabilistic Prediction of Yangtze River Valley Summer Rainfall.” Advances in Atmospheric Sciences 32: 497–504. doi:10.1007/s00376-014-4073-8.

Liu, J. G., and Z. H. Xie. 2013. “Improving Simulation of Soil Moisture in China Using a Multiple Meteorological Forcing Ensemble Approach.” Hydrology and Earth System Sciences 17: 3355–3369. doi:10.5194/hess-17-3355-2013.

Liu, J. G., and Z. H. Xie. 2014. “BMA Probabilistic Quantitative Precipitation Forecasting over the Hualie Basin Using TIGGE Multi-Model Ensemble Fore- Casts.” Monthly Weather Review 142: 1542–1555. doi:10.1175/MWR-D-13-0031.1,2014.

Niu, G. Y., Z. L. Yang, R. E. Dickinson, and L. E. Gulden. 2005. “A Simple TOPMODEL- Based Runoff Parameterization (SIMTOP) for Use in Global Climate Models.” Journal Geophys Researcher 110: D21106. doi:10.1029/2005JD006111.
Niu, G. Y., Z. L. Yang, R. E. Dickinson, L. E. Gulden, and H. Su. 2007. "Development of a Simple Groundwater Model for Use in Climate Models and Evaluation with Gravity Recovery and Climate Experiment Data." *Journal Geophys Researcher* 112: D07103. doi:10.1029/2006JD007522.

Oleson, K. W., D. M. Lawrence, G. B. Bonan, B. Drevniak, M. Huang, C. D. Koven, S. Levis, et al. 2013. "Technical Description of Version 4.5 Of the Community Land Model (CLM)." In *NCAR Technical Note NCAR/TN-503+STR*, 422. Boulder, CO: National Center for Atmospheric Research. doi:10.5065/D6RR1W7M.

Oleson, K. W., G. Y. Niu, Z. L. Yang, D. M. Lawrence, P. E. Thornton, P. J. Lawrence, R. Stockli, R. E. Dickinson, G. B. Bonan, and S. Levis. 2007. "CLM 3.5 Documentation." 34 pp. http://www.cgd.ucar.edu/tss/clm/distribution/clm3.5/CLM3_5_documentation.pdf

Philippon, N., and B. Fontaine. 2002. "The Relationship between the Sahelian and Previous 2nd Guinean Rainy Seasons: A Monsoon Regulation by Soil Wetness?." *Annales Geophysicae* 20: 575–582. doi:10.5194/angeo-20-575-2002.

Qian, T., A. Dai, K. E. Trenberth, and K. W. Oleson. 2006. "Simulation of Global Land Surface Conditions from 1948 to 2004. Part I: Forcing Data and Evaluations." *Journal of Hydrometeorology* 7: 953–975. doi:10.1175/JHM540.1.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski. 2005. "Using Bayesian Model Averaging to Calibrate Forecast Ensembles." *Monthly Weather Review* 133: 1155–1174. doi:10.1175/MWR2906.1.

Rowlands, D. D., S. B. Luthcke, S. M. Klosko, F. G. R. Lemoine, D. S. Chinn, J. J. McCarthy, C. M. Cox, and O. B. Anderson. 2005. "Resolving Mass Flux at High Spatial and Temporal Resolution Using GRACE Intersatellite Measurements." *Geophys Research Letters* 32: L04310. doi:10.1029/2004GL021908.

Sheffield, J., G. Goteti, and E. F. Wood. 2006. "Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling." *Journal of Climate* 19: 3088–3111. doi:10.1175/JCLI3790.1.

Sun, Q., Z. H. Xie, and X. J. Tian. 2015. "GRACE Terrestrial Water Storage Data Assimilation Based on the Ensemble Four-Dimensional Variational Method PODEn4DVar: Method and Validation." *Science China: Earth Sciences* 58: 371–384. doi:10.1007/s11430-014-4978-1.

Swenson, S., and J. Wahr. 2006. "Post-Processing Removal of Correlated Errors in GRACE Data." *Geophys Researcher* 33: L08402. doi:10.1029/2005GL025285.

Swenson, S. C., and D. M. Lawrence. 2012. "A New Fractional Snow Covered Area Parameterization for the Community Land Model and Its Effect on the Surface Energy Balance." *Journal Geophys Researcher* 117: D21107. doi:10.1029/2012JD018178.

Syed, T. H., J. S. Famiglietti, M. Rodell, J. L. Chen, and C. R. Wilson. 2008. "Analysis of Terrestrial Water Storage Changes from GRACE and GLDAS." *Water Resources Researcher* 44: 1–15. doi:10.1029/2006WR005779.

Taylor, K. E. 2001. "Summarizing Multiple Aspects of Model Performance in a Single Diagram." *Journal of Geophysical Research* 106: 7183–7192. doi:10.1029/2000JD900719.

Tian, X., Z. Xie, A. Dai, B. Jia, and C. Shi. 2010. "A Microwave Land Data Assimilation System: Scheme and Preliminary Evaluation over China." *Journal Geophys Researcher* 115: D21113. doi:10.1029/2010JD014370.

Wahr, J., S. Swenson, V. Zlotnicki, and I. Velicogna. 2004. "Time Variable Gravity from GRACE: First Results." *Geophys Research Letters* 31: L11501. doi:10.1029/2004GL019779.

Wang, A., D. P. Lettenmaier, and J. Sheffield. 2011. "Soil Moisture Drought in China, 1950–2006." *Journal of Climate* 24: 3257–3271. doi:10.1175/2011JCLI3733.1.

Wang, A., and X. Zeng. 2011. "Sensitivities of Terrestrial Water Cycle Simulations to the Variations of Precipitation and Air Temperature in China." *Journal Geophys Researcher* 116: D02107. doi:10.1029/2010JD014659.

Wang, K. C., and R. E. Dickinson. 2012. "A Review of Global Terrestrial Evapotranspiration: Observation, Modeling, Climatology, and Climatic Variability." *Reviews of Geophysics* 50: RG2005. doi:10.1029/2011RG000373.

Zeng, N., J. H. Yoon, M. Annarita, and S. Sean. 2007. "Variability of Basin Scale Terrestrial Water Storage from a PER Water Budget Method: The Amazon and the Mississippi." *Journal of Climate* 21: 248–264. doi:10.1175/2007JCLI1639.1.