Journal of Hydrology 596 (2021) 125730

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

High-resolution spatially explicit land surface model calibration using field-scale satellite-based daily evapotranspiration product

Yi Yang [a,*], Kaiyu Guan [a,b,c,*], Bin Peng [a,b], Ming Pan [d], Chongya Jiang [a,c], Trenton E. Franz [e]

[a] College of Agricultural, Consumer and Environmental Sciences, University of Illinois at Urbana Champaign, Urbana, IL, USA
[b] National Center for Supercomputing Applications, University of Illinois at Urbana Champaign, Urbana, IL, USA
[c] Center for Advanced Bioenergy and Bioproducts Innovation, University of Illinois at Urbana Champaign, Urbana, IL, USA
[d] Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ, USA
[e] School of Natural Resources, University of Nebraska-Lincoln, Lincoln, NE, USA

ARTICLE INFO

This manuscript was handled by G. Syme, Editor-in-Chief, with the assistance of Ashok Mishra, Associate Editor

Keywords:
High-resolution modeling
Evapotranspiration
Land surface model
BESS
STAIR
Parameter estimation

ABSTRACT

High-resolution simulation of water budgets across the agricultural landscape is critically important to a variety of applications, such as precision agriculture, water resources management, and environmental quality assessment. Model-data integration has been shown to be an effective approach to reduce model uncertainties and there is a growing opportunity to improve land surface modeling through spatially explicit calibration with satellite data in recent decades. Recently, a satellite-based daily 30-m resolution evapotranspiration (ET) product BESS-STAIR has been developed, achieving a high performance and well capturing the spatial and temporal dynamics of ET across the U.S. Corn Belt. To explore the potential of high-resolution spatially explicit calibration for advancing land surface modeling at fine scales, we carried out calibration experiments for the Noah-MP land surface model (LSM) over cropland using this newly developed BESS-STAIR ET. We first used Sobol sensitivity analysis to identify the most sensitive parameters for the Noah-MP's ET simulation. The most sensitive vegetation (minimum stomatal resistance) and soil parameters (saturated hydraulic conductivity, saturated matric potential, and a soil pore size distribution parameter) were calibrated using BESS-STAIR ET to improve model simulation of surface water balance. We conducted calibration experiments at 8 eddy covariance flux tower sites that grew maize and soybean across the U.S. Corn Belt, as well as a regional calibration study on the Spoon River watershed in Champaign, Illinois. When benchmarked with flux tower measurements, the BESS-STAIR ET–calibrated model (driven by flux tower forcing) on average reduced the RMSE of hourly ET from 61 W/m^2 to 47 W/m^2 for maize, and from 66 W/m^2 to 53 W/m^2 for soybean, and matched the performance of directly calibrating using flux tower measured ET. The regional study found that calibration using BESS-STAIR ET also improved the simulation of long-term regional water budgets and achieved better performance of ET than traditionally lumped calibration using streamflow. Further analysis revealed that the high-resolution calibration can resolve the spatial variations of ET to a certain extent, and the accuracy of the calibration can be largely attributed to the low bias and excellent long-term correlation of the BESS-STAIR ET data itself. Our study thus demonstrates the effectiveness of high-resolution model calibration and provides important implications in field-scale hydrological modeling and precision agricultural applications.

1. Introduction

High-resolution simulation of surface energy and water budgets can benefit a variety of purposes (Wood and Coauthors, 2011). Field scale modeling across the agricultural landscape in particular is of great interest for precision agricultural applications (Karthikeyan et al., 2020; Lobell et al., 2015; Wood and Coauthors, 2011). For instance, simulations that resolve water and energy budgets in individual crop fields can be used to study crops' response to various environmental factors and guide producers' management practices like irrigation. Furthermore, resolving water-energy budgets is a prerequisite to simulate nutrient dynamics (Austin et al., 2004; Ciais et al., 2013), which may provide implications for fertilization management.

A major source of uncertainty within land surface modeling lies in

* Corresponding authors at: College of Agricultural, Consumer and Environmental Sciences, University of Illinois at Urbana Champaign, Urbana, IL, USA.
E-mail addresses: yiy12@illinois.edu (Y. Yang), kaiyug@illinois.edu (K. Guan).

https://doi.org/10.1016/j.jhydrol.2020.125730
Received 3 September 2020; Received in revised form 24 October 2020; Accepted 1 November 2020
Available online 6 November 2020

the parameters related with surface energy partitioning, canopy dynamics and hydrological processes (Crow et al., 2003; Xia et al., 2002). This is particularly true at high resolution due to the spatial heterogeneity of parameters. Certain parameters, like those related with plant responses to abiotic stresses (Jarvis, 1976), are inherently empirical and hard to measure directly. Other parameters, like soil hydraulic properties, are physically based and can be measured, albeit at small support volumes from soil cores (~250 cm^3). However, they are highly heterogeneous in space and still have high uncertainties due to limited sampling coverage. As a result, the successful application of land surface models, to a large extent, is contingent upon how well the parameters are calibrated.

Traditionally, calibration of hydrological models is conducted only at basin scales using streamflow observations pending on the availability of a stream gauge (Immerzeel and Droogers, 2008; Rajib et al., 2018; Sutanudjaja et al., 2014). This kind of calibration work usually results in an apparently accurate model, giving acceptable stream simulation while unrealistically representing internal hydrological processes in individual model grids. In addition, the equifinality issue (Beven, 2006) further complicates the model calibration problem as the degree of equifinality will significantly increase when observational constraints are limited. While calibration using stream gauges at multiple locations can slightly mitigate the equifinality problem and improve the overall performance (Niraula et al., 2012), the internal hydrological processes still cannot be fully represented. Parameter regionalization techniques can be used to estimate hydrologic parameters in ungauged areas even across scales (Beck et al., 2016), but transferring parameters to finer resolutions is challenging and the modeling performance with regionalized parameters cannot match direct calibration (Beck et al., 2016; Samaniego et al., 2010). Moreover, many hydrological variables (e.g., evapotranspiration) and parameters (vegetation and soil parameters) have large spatial heterogeneities at fine scales that traditional lumped calibration is unable to capture. Therefore, it is still appealing to conduct spatially explicit calibration using distributed observations especially at high resolutions (Yang et al., 2019).

The advancement in satellite remote sensing has enabled hydrologists and environmental modelers to use remotely sensed measurements of water and energy balance components to constrain land surface models in a spatially explicit way. Remotely sensed soil moisture from passive microwave satellite is widely used to calibrate soil hydraulic parameters and improve model simulations (Houser et al., 1998; Kerr et al., 2001; Wanders et al., 2014). However, due to coarse resolutions (tens of kilometers), shallow penetration depths (Jackson et al., 1997; Peng et al., 2017), and high uncertainties of microwave-based satellite soil moisture products, it is still challenging to calibrate a field-scale model with satellite-based soil moisture estimations over an extensive area. Satellite remote sensing can also provide estimation for evapotranspiration (ET), a critical flux that connects surface energy and water budgets. Compared with soil moisture products, ET estimations from satellite can have a much higher spatial resolution and better accuracy (Anderson et al., 2011; Li et al., 2009; Liou and Kar, 2014; Su, 2002). Previous studies have explored the potential of using satellite-based ET products for hydrological model calibration (Immerzeel and Droogers, 2008; Rajib et al., 2018), but they mainly focused on using MODIS ET product, which has been demonstrated to bear large uncertainties (Velpuri et al., 2013). Further, very few of those studies focused on agricultural landscapes, where field-scale water management is crucial for crop growth and yield. Recently, a satellite-based daily, cloud-/gap-free, 30-m resolution ET product BESS-STAIR has been developed (Jiang et al., 2019). Validations showed that the carefully gap-filled satellite data and physically based retrieval approach, which considers multiple constraints (energy, water, and carbon) in the ET process (Jiang et al., 2019), provide an extremely powerful and reliable way to reconstruct ET. This is evidenced by BESS-STAIR's good accuracy with an overall R^2 of 0.75 and root mean squared error (RMSE) of 24.3 W/m^2 when benchmarked with daily eddy-covariance measurements at 12 flux tower sites across the U.S. Corn Belt. It also well captured the spatial and temporal patterns of ET. BESS-STAIR ET provides a great opportunity to test whether and how satellite products could improve the performance of land surface models at fine scales.

The primary objective of this paper is to investigate the effectiveness of high-resolution spatially explicit land surface model calibration using BESS-STAIR ET product and the Noah-MP land surface model (LSM). We aim to answer the following three science questions in this study: (1) To what extent can calibration using the high-resolution ET data improve Noah-MP's simulation of surface water budgets? (2) What are the benefits of spatially explicit model calibration at fine scales compared to traditional lumped calibration? (3) What factors determine the effectiveness of the high-resolution model calibration experiment (and their implications)? To answer these questions, we conducted model calibration experiments to constrain the most sensitive vegetation and soil parameters in Noah-MP using BESS-STAIR ET at 8 Ameriflux sites as well as a regional calibration study in the Spoon River Watershed in Champaign, Illinois. We evaluated the accuracy of the ET simulation from calibrated models using flux tower measurements at the site level, and its spatial patterns and comparison with traditional lumped calibration in the regional study. Simulations of other surface water budget components (soil water content and runoff) were also evaluated. Finally, the effectiveness of the high-resolution land surface model calibration and its implications were discussed.

2. Materials and methods

2.1. Noah-MP LSM and related parameters

In this study, we chose the widely used Noah-MP LSM for its relatively complete and parsimonious representation of surface water and energy budgets as well as its high computational efficiency. Noah-MP is an improved version of the original Noah model (Niu et al., 2011; Yang et al., 2011) that enhances the representation of physics and adds multi-parameterization options. In this study, we ran Noah-MP in offline mode without coupling to atmospheric models. We used default parameterization options except for stomatal resistance, for which we used the Javis scheme (Jarvis, 1976). As the primary goal of this study is to investigate the effectiveness of high-resolution calibration using BESS-STAIR ET to improve Noah-MP's simulation of water and energy balances, we reduced the model complexity by turning off carbon cycling of Noah-MP; thus, Javis stomatal resistance scheme is deemed adequate and can be better constrained for this study.

In Noah-MP, ET is resolved as the sum of the three components: canopy evaporation, bare ground evaporation, and transpiration. During the growing season, transpiration dominates ET, so we focused on parameterization schemes and parameters that control the transpiration process. In Noah-MP, the transpiration heat flux TR_v (W/m^2) is calculated as:

$$TR_v = f_{veg} \cdot \rho_{air} \cdot C_{p,air} \cdot C_{tW} \cdot (e_{sat,tv} - e_{a,H})/\gamma \quad (1)$$

where f_{veg} is the fraction of the ground covered by vegetation, ρ_{air} is the density of air (g cm^{-1}), $C_{p,air}$ is the heat capacity (J kg^{-1} K^{-1}) of dry air at constant pressure, $e_{sat,v}$ is the saturation vapor pressure (Pa) inside the leaf at leaf temperature tv (K), $e_{a,H}$ is the canopy air vapor pressure (Pa) and γ is the psychrometric constant (Pa K^{-1}). C_{tW} is the transpiration conductance (m s^{-1}) from leaf to canopy air which represents the plant physiological and environmental controls on canopy transpiration C_{tW} is given as:

$$C_{tW} = (1 - f_{wet}) \cdot (LAI_{sun}/(r_b + r_{s,sun}) + LAI_{sha}/(r_b + r_{s,sha})) \quad (2)$$

where f_{wet} is the wetted fraction of canopy, LAI is the leaf area index, r_b is the bulk leaf boundary layer resistance (s m^{-1}), r_s is the leaf stomatal resistance (s m^{-1}), and subscripts *sun* and *sha* denote the fraction of sunlit and shaded leaves, respectively.

In Jarvis scheme, stomatal resistance r_s is the minimum stomatal resistance (s m^{-1}) scaled by the effects of radiation, air temperature, vapor pressure deficit and soil water stress:

$$r_s = r_{s,min}/(r_{c,s} \cdot r_{c,T} \cdot r_{c,Q} \cdot r_{c,soil}) \tag{3}$$

where $r_{s,min}$ is the minimum stomatal resistance which corresponds to the highest possible transpiration rate. The four variables $r_{c,s}$, $r_{c,T}$, $r_{c,Q}$ and $r_{c,soil}$ are scalers of solar radiation, air temperature, vapor pressure deficit and soil water stress, respectively, reflecting the environmental controls on plant stomatal dynamics. These scalers are given as:

$$r_{c,s} = \frac{\left(\frac{2PAR}{RGL} + \frac{r_{s,min}}{r_{s,max}}\right)}{1 + \frac{2PAR}{RGL}} \tag{4}$$

$$r_{c,T} = 1 - 0.0016\left(T_{opt} - T_{sfc}\right) \tag{5}$$

$$r_{c,Q} = \frac{1}{1 + HS + max(q_{2,sat} - q_2, 0)} \tag{6}$$

$$\mathrm{r_{c,soil}} = \sum_{\mathrm{i=1}}^{\mathrm{n_{root}}} \frac{(\theta_\mathrm{i} - \theta_\mathrm{w})\mathrm{d_{z_i}}}{(\theta_\mathrm{ref} - \theta_\mathrm{w})\sum_{\mathrm{j=1}}^{\mathrm{n_{root}}} \mathrm{d_{z_i}}} \tag{7}$$

where RGL is the radiation stress parameter, T_{opt} is the optimal temperature for transpiration (K), HS is the VPD stress parameter, $r_{s,max}$ is the maximum stomatal resistance (s m^{-1}), θ_i represents the volumetric soil water content (SWC; m^3 m^{-3}) in the ith layer, θ_{ref}, and θ_w are the reference soil water content (m^3 m^{-3}) and soil water content at wilting point (m^3 m^{-3}), and d_{z_i} is the ith soil layer depth (m).

In Noah-MP, soil moisture exerts its control on ET through the soil water stress term. Aside from the two soil parameters θ_{ref} and θ_w that are directly involved in soil water stress calculation in Eq. (7), the soil water content in each layer also determines the term. Noah-MP has four soil layers with thicknesses of 0.1 m, 0.3 m, 0.6 m, and 1 m from top to bottom. It solves the 1-dimensional Richards equation to calculate the soil water content of each layer and the flux between them:

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial z}\left[K\left(\frac{\partial\psi}{\partial\theta}\right)\frac{\partial\theta}{\partial x}\right] + \frac{\partial K}{\partial z} + F_\theta \tag{8}$$

where θ is the volumetric soil water content (m^3 m^{-3}), t is time (s), z is depth (m), F_θ is the source term (s^{-1}). K and ψ are the unsaturated soil hydraulic conductivity (m s^{-1}) and matric potential (mH_2O), which can be calculated using the corresponding saturated values and θ using the Campbell model:

$$\psi = \psi_{sat}\left(\frac{\theta}{\theta_{max}}\right)^{-B} \tag{9}$$

$$K = K_{sat}\left(\frac{\theta}{\theta_{max}}\right)^{2B+3} \tag{10}$$

The parameters involved in (9) and (10) are: saturated soil hydraulic conductivity K_{sat} (m s^{-1}), saturated soil matric potential ψ_{sat} (mH_2O), porosity θ_{max} (m^3 m^{-3}) and the B parameter, which reflects the pore size distribution of soil particles. Additionally, Noah-MP uses the four parameters above to derive reference soil water content θ_{ref} and soil water content at wilting point θ_w:

$$\theta_{ref1} = \theta_{max} \cdot \left(\frac{5.79e-9}{K_{sat}}\right)^{\frac{1}{2\cdot B+3}} \tag{11}$$

$$\theta_{ref} = \theta_{ref1} + \frac{\theta_{max} - \theta_{ref1}}{3} \tag{12}$$

$$\theta_{wilt} = 0.5 \cdot \theta_{max} \cdot \left(\frac{200}{\psi_{sat}}\right)^{\frac{-1}{B}} \tag{13}$$

Noah-MP makes no distinction of soil properties between different layers, so the six soil parameters are the same at different depths.

A total of nine relevant parameters to ET calculation are summarized in Table 1. Default parameter values are needed to provide a prior distribution for calibration and a performance baseline. In this study, the default values for the five vegetation parameters ($r_{s,\mathrm{min}}$, $r_{s,\mathrm{max}}$, RGL, HS, and T_{opt}) were acquired from the cropland category in the default model parameter lookup table. Three soil parameters (K_{sat}, θ_{max}, B) were acquired from the Probabilistic Remapping of SSURGO (POLARIS) dataset (Chaney et al., 2016). We calculated the weighted average from the six layers in POLARIS (with thicknesses of 0.05 m, 0.1 m, 0.15 m, 0.3 m, 0.4 m, 1 m) to match the four layers defined in Noah-MP. The saturated soil matric potential ψ_{sat} was acquired from Noah-MP's lookup table according to the dominant soil type provided by the SSURGO soil texture database. Runoff in the Noah-MP model is calculated as the sum of surface runoff and subsurface runoff (runoff and groundwater parameterization option 3). Surface runoff is a combination of saturation excess and infiltration excess runoff. In the regional calibration study with runoff as one of the constraints, we additionally calibrated two parameters: slope parameter controlling subsurface runoff and surface runoff parameter (REFKDT; (Schaake et al., 1996)) controlling surface runoff. Since we focused on evaluating the effectiveness of calibration using satellite-based ET data, we used site-measured or satellite-derived leaf area index (LAI) as model input instead of using lookup table-based default LAI to reduce uncertainties and simplify the problem.

2.2. Data

2.2.1. Site data

We chose eight AmeriFlux sites (Table 2) across the U.S. Corn Belt for this study. Either maize or soybean was grown at these sites, representing major crops in the U.S. Corn Belt. US-Ne1 and US-Ne2 are irrigated, while others are rainfed. Eddy covariance systems were installed at each site to observe energy, water and carbon fluxes.

Six meteorological forcings are required to drive Noah-MP, including surface air temperature (K), relative humidity, downward shortwave radiation (W/m^2), downward longwave radiation (W/m^2), precipitation (kg m^{-2} s^{-1}) and wind speed (m s^{-1}). We acquired data from the FLUXNET2015 database for US-Ne1, US-Ne2 and US-Ne3, and from the AmeriFlux database for other sites. The data from FLUXNET2015 was gap-filled. We used ERA-interim data to fill data gaps at other AmeriFlux sites following a similar approach used for FLUXNET2015 (Vuichard and Papale, 2015). All data were aggregated to hourly interval for consistency. We refer to these forcing as site forcing (SF) in the remainder of this paper.

LAI is the key vegetational forcing to drive Noah-MP. We used linearly interpolated field-measured LAI from the Carbon Sequestration Program at UNL ARDC (http://csp.unl.edu/Public/sites.htm) for US-Ne1, US-Ne2 and US-Ne3, and satellite-derived LAI for other sites. Satellite LAI was derived from daily 30-m resolution surface reflectance data fused from MODIS and Landsat satellites (Luo et al., 2018). A statistical model calibrated using field-measured LAI was employed for satellite LAI estimation (Jiang et al., 2019).

Site measured latent heat flux (LE; W/m^2) was used as the benchmark for Noah-MP simulation. Similar to meteorological forcings, we acquired LE data from the FLUXNET2015 database for US-Ne1, US-Ne2 and US-Ne3, and from the AmeriFlux database for other sites. We did not apply gap-filling and only available observation data were used. All data were aggregated to hourly interval for consistency.

2.2.2. BESS-STAIR ET data

We investigated the effectiveness of using BESS-STAIR ET in improving Noah-MP. BESS is a satellite-driven biophysical model to monitor water and carbon fluxes (Jiang and Ryu, 2016; Ryu et al., 2011). It considers multiple physical constraints of ET process with respect to energy, water, and carbon, and uses environmental and vegetation information derived from multi-satellite data as inputs to

Table 1
Parameters related with ET calculation in the Noah-MP land surface model. Subscript 0 denotes the default value of the corresponding parameter.

Parameter	Units	Range	Default	Description	Range in calibration
$r_{s,min}$	sm^{-1}	10–1000	40	minimum stomatal resistance	10–1000
$r_{s,max}$	sm^{-1}	2000–10000	5000	maximum stomatal resistance	–
RGL	–	30–100	100	radiation stress parameter	–
HS	–	36–47	36.25	vapor pressure deficit stress parameter	–
T_{opt}	K	293–303	298	optimum temperature for transpiration	–
θ_{ref}	–	0.171–0.37	–	reference soil water content	–
θ_{wilt}	–	0.01–0.17	–	soil water content at wilting point	–
B	–	2–13	–	the B parameter	$0.8B_0 - 1.35B_0$
K_{sat}	ms^{-1}	$9.55 \times 10^{-7} - 5.01 \times 10^{-5}$	–	saturated soil hydraulic conductivity	$max({K_{sat0}}^{1.15}, 10^{-4.4}) - min({K_{sat0}}^{0.85}, 10^{-6.2})$
θ_{max}		0.371–0.48	–	maximum soil water content	
ψ_{sat}	m	0.03–0.76	–	saturated soil matric potential	$max(0.65\psi_{sat0}, 0.2) - min(1.35\psi_{sat0}, 0.85)$

Table 2
Information of AmeriFlux sites used in this study.

Site	Year	Longitude	Latitude	Elevation	Mean precipitation/ mm	Mean temperature/°C	Rainfed/ Irrigated	Crop type
US-Bo1	2000–2008	40.0062	−88.2904	219	792.10	11.40	rainfed	Maize in odd years soybean in even years
US-Br1	2005–2011	41.9749	−93.6906	313	933.61	9.13	rainfed	odd maize, even soybean
US-IB1	2005–2011	41.8593	−88.2227	227	966.76	9.52	rainfed	even maize, odd soybean
US-Ro1	2004–2012	44.7143	−93.0898	260	762.45	7.71	rainfed	odd maize, even soybean
US-Ne1	2003–2012	41.1651	−96.4766	361	840.40	10.60	irrigated	continuous maize
US-Ne2	2003–2012	41.1649	−96.4701	362	871.80	10.32	irrigated	odd maize even soybean before 2009, maize beginning 2009
US-Ne3	2003–2012	41.1797	−96.4397	363	712.15	10.42	rainfed	odd maize, even soybean
US-Br3	2005–2011	41.9747	−93.6936	313	836.91	9.22	rainfed	even maize, odd soybean

estimate ET (Jiang and Ryu, 2016). STAIR is an algorithm fusing high spatial resolution Landsat data and high temporal resolution MODIS data to generate daily 30 m resolution surface spectral reflectance (Luo et al., 2018), which was used as inputs of BESS (Jiang et al., 2019). By integrating BESS ET model and STAIR fusion data, the BESS-STAIR framework provides accurate cropland ET estimations at 30-m resolution and daily interval.

2.2.3. Data for regional calibration

Aside from calibration experiments at site level using flux tower forcing, we did additional experiments using regionally available forcing both at site level and in a region to further assess the benefits of BESS-STAIR ET for calibration. Princeton University Meteorological Forcing Dataset (PUMET) provides all the meteorological variables required to run Noah-MP LSM at 4-km spatial resolution and hourly time scale (Beck et al., 2019; Pan et al., 2016). PUMET integrates reanalysis data with observations and uses downscaling to achieve higher spatial and temporal resolution. In addition to estimating ET, BESS has also been used to estimate downward shortwave radiation at 5 km resolution using MODIS atmospheric and land products (Ryu et al., 2018), which achieved better performance than downward shortwave radiation in PUMET meteorological forcing dataset (See Fig. S1). In this study, we thus used downward shortwave radiation from this BESS-derived product and other forcing variables from PUMET as the regional forcing (RF). Since US-Ne1 and US-Ne2 are irrigated sites, and the irrigation amounts have been added to precipitation in FLUXNET2015 dataset, precipitation of these two sites was still from site forcing. In the regional study, the precipitation in the PUMET forcing was bias-corrected using the observed precipitation data from the NOAA US1ILCP0031 weather station located in the study region. Crop types in each individual field were obtained from USDA Cropland Data Layer (CDL). The streamflow data at the watershed outlet for lumped calibration and evaluation was from the USGS 03,336,890 SPOON RIVER station.

2.3. Model simulations, sensitivity analysis and calibration experiment

The following experiments were conducted in this study: default model simulations, sensitivity analysis, model calibration using flux tower ET or BESS-STAIR ET and a regional calibration study in the Spoon River watershed (Fig. 1 (a)).

2.3.1. Default model simulation

To provide a performance baseline before calibration, we ran Noah-MP with default model parameters, i.e., cropland category in the lookup table for vegetation parameters, POLARIS combined with SSURGO for soil parameters. Before each simulation, one-year spin-up was conducted, which is sufficient for simulating soil water dynamics in Noah-MP (Gutmann and Small, 2010). We ran default Noah-MP simulation with site forcing, the original PUMET and PUMET + BESS-derived downward shortwave radiation (regional forcing). The RMSE, bias and coefficient of determination (R^2) were calculated between the simulated hourly ET and flux tower measurements from the peak growing season (Jun 15 – Sep 14). When comparing the performance of model simulations driven by different forcings, we omitted US-Ne1 and US-Ne2 because the precipitation directly from PUMET is unrealistic in the two irrigated sites. The performance of default model simulations using different forcings are provided in the supplementary materials.

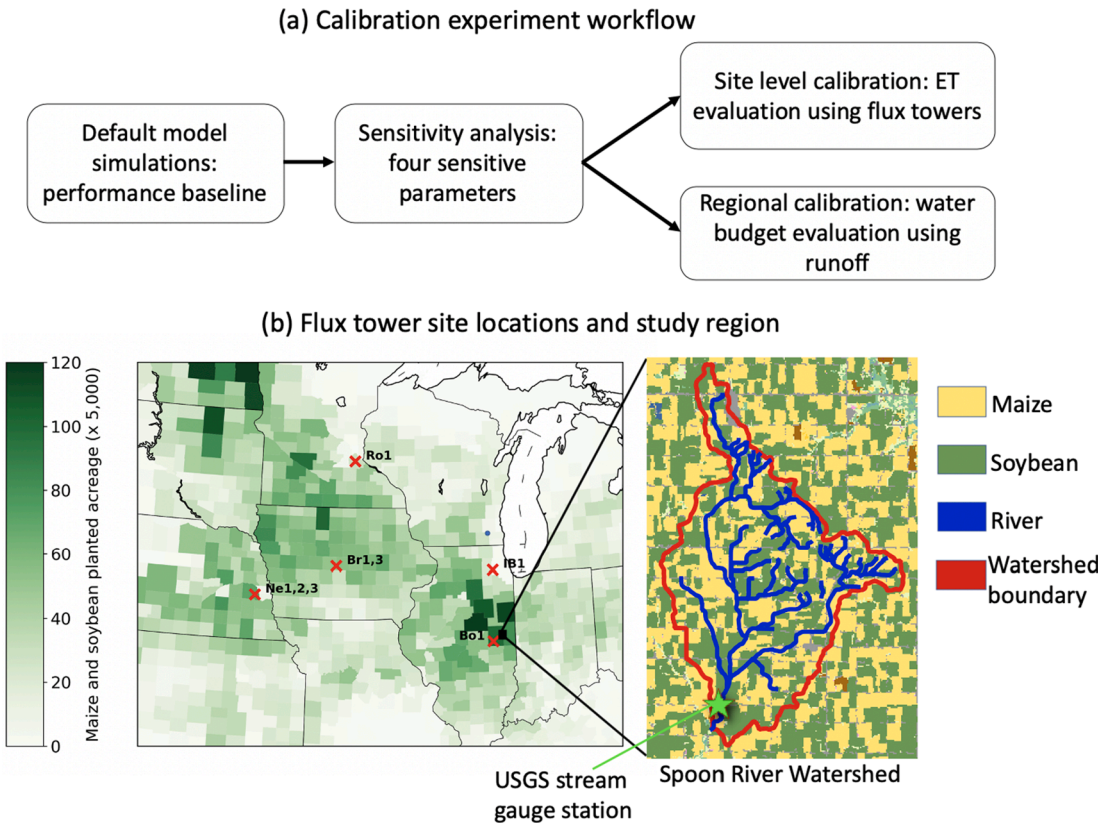

Fig. 1. (a) Calibration experiment workflow; and (b) study sites and region. The Midwest map shows the county-level planted acreage of maize and soybean (data from USDA). The background layer of the study region map is CDL data in 2014.

2.3.2. Parameter sensitivity analysis

To identify the most sensitive parameters for calibration experiments, we conducted Sobo sensitivity analysis (Saltelli et al., 2010; Sobol, 2001) to determine the sensitivities of the 9 parameters related to ET calculation in Noah-MP (Table 1). Sobol's method is a variance based global sensitivity analysis method that can determine the contributions from both the individual parameters and the interactions between them. To calculate the first-order and total-order sensitivity indices, we needed to sample $N\times(d+2)$ parameter sets, where N is the number of samples generated each time, d is the parameter space dimension. The parameter sets were sampled from the Sobol quasi-random sequence following the Monte Carlo scheme with two ensembles of size N first sampled and then cross sampling performed by holding one parameter fixed at a time. In our case, $d=9$ and we set $N=300$ resulting in 3300 parameter sets in total. The ranges of the parameters were shown in Table 1. Log space sampling was used for parameters covering 2 or more orders of magnitude. We then ran the Noah-MP model continuously through all the available data at each site on all parameter sets. The first-order and total-order Sobol sensitivity indices were calculated based on the RMSEs between Noah-MP simulated hourly ET and flux tower measurements.

2.3.3. Site-level calibration experiments

With the most sensitive parameters identified, we first calibrated the model directly using flux tower measured ET to test different calibration schemes. We calibrated the chosen parameters by minimizing the RMSE between model simulated hourly ET and flux tower measurements. We used the University of Arizona shuffled complex evolution (SCE-UA) algorithm to search the parameter space with 200–400 iterations. SCE-UA is a powerful and robust global optimization algorithm (Duan et al., 1993) and has been widely used for parameter calibration in hydrological modeling (Duan et al., 1992; Sorooshian et al., 1993). The default parameter values were used as the initial guess in the SCE-UA algorithm. All metrics for both calibration and evaluation were calculated using the data from the peak growing season.

We first tested two calibration schemes using site ET to assess the calibration and validation performance: (1) year-by-year calibration, and (2) multi-year calibration. Year-by-year calibration used ET observations in each growing season to constrain the parameters for any specific year and gives the "calibration" performance separately for each year. Multi-year calibration used leave-one-year-out validation method, i.e. reserving one year for validation, while using all other years with available data when the same crop was planted for parameter calibration. Multi-year calibration provides the "validation" performance. We compared the performance of multi-year calibration and year-by-year calibration to evaluate the "validation" performance and "calibration" performance. Both multi-year and year-by-year calibration experiments were implemented using a two-step approach. In the first step, soil and vegetation parameters were calibrated simultaneously. The median values of the calibrated soil parameters in the first step were directly used in the second step experiments, in which we only calibrated sensitive vegetation parameters. This two-step approach was reasonable as we would not expect significant interannual variations in soil hydraulic properties, or we just could not capture these variations through model parameter calibration if there is any. Final calibration performance was reported using the second step results in both the multi-year and year-by-year calibration experiments.

At each site, BESS-STAIR ET data was then used to calibrate Noah-MP following the same procedures as using site ET. Since BESS-STAIR only provides daily ET instead of hourly ET, the objective function was changed to the RMSE of daily ET between Noah-MP simulation and BESS-SATIR estimation. There was little difference between calibration using hourly ET or daily ET (the RMSE on average was 2.5 W/m^2 lower for calibration using hourly site ET than using daily site ET at Ne1, Ne2 and Ne3 when the performance was evaluated using hourly site ET), so

the time scale difference between daily and hourly for calibration should not be a major source of difference. We evaluated the performance of default and calibrated models in simulating surface energy budgets for maize and soybean separately using three statistical metrics: RMSE, R^2 and bias. We also analyzed the results of calibrated parameters and compared the difference between calibration using site ET and BESS-STAIR ET. The simulated first layer SWC before and after calibration at Ne1, Ne2 and Ne3 was evaluated using in-situ Time Domain Reflectometry (TDR) sensor measurements (Ledieu et al., 1986). In addition to RMSE, R^2 and bias, unbiased RMSE (ubRMSE) was also calculated for SWC evaluation.

2.3.4. Regional calibration

We then conducted a regional calibration study on the Spoon river watershed, located east of Champaign county, Illinois (Fig. 1 (b)) The regional calibration study was intended to (1) evaluate the performance of the high-resolution calibration in terms of spatial patterns; (2) evaluate the performance of the simulation of regional water budgets (runoff); and (3) provide a concrete example of applying the high-resolution calibration using BESS-STAIR ET data at regional scales. In the watershed, field boundaries were extracted by combining the Common Land Unit (CLU) and multi-year CDL data, both from USDA, resulting in a total of 1215 fields. The calibration was done at each individual field instead of individual pixels to reduce the computational cost. The regional forcing data described in 2.2.3 was used to drive the model and the BESS-STAIR ET data was aggregated to field level as the reference. After the calibration, each field has its own calibrated soil parameters and two sets of vegetation parameters for maize and soybean respectively. We calibrated the model from 2004 to 2007 and the evaluation was done from 2014 to 2017 from two aspects: first, the model simulated ET was evaluated again using BESS-STAIR ET across the region and the performance metrics were calculated, which allows us to examine the spatial patterns of the simulation performance; second, we evaluated the model simulated regional water budget using the streamflow data at the watershed outlet. To evaluate the benefits of the spatially explicit calibration using high-resolution ET, we further compared the performance of (1) traditional lumped calibration using streamflow only and (2) joint streamflow and ET calibration with ET calibrated in a spatially explicit manner. In the lumped calibration, the initial parameter values were different in different fields (POLARIS for soil parameters) but they moved in the same direction and the same relative magnitude during calibration. The calibration and validation configurations are the same as before.

3. Results

3.1. Parameter sensitivity analysis

The first and total order sensitivity indices for related parameters are shown in Fig. 2. Minimum stomatal resistance had the highest sensitivity in general while the other 4 vegetation parameters showed marginal effects. Three soil parameters, i.e., K_{sat}, B and ψ_{sat} also showed significant sensitivities, while other soil parameters had non-significant effects. Therefore, we identified the 4 most sensitive parameters as $r_{s,min}$, K_{sat}, B, and ψ_{sat}. We only chose the most sensitive 4 parameters because the sensitivities of the remaining parameters were significantly smaller (Fig. 2). We only calibrated those 4 parameters in the calibration experiments.

3.2. Site level calibration using flux tower and BESS-STAIR ET

3.2.1. Comparison of calibration and validation performance using flux tower ET

With site forcing, both multi-year and year-by-year calibration significantly improved the performance of simulated ET. The performance differences between these two calibration schemes were marginal, which means the calibration and validation performances are similar (Fig. 3). Overall, multi-year and year-by-year calibrations improved the model performance in simulating ET. For maize, RMSEs of ET simulation decreased from 62 W/m^2 in default simulation to 46 and 45 W/m^2 in multi-year and year-by-year calibrations, respectively, and for soybean, RMSEs of ET simulation decreased from 66 W/m^2 in default simulation to 52 and 48 W/m^2 in multi-year and year-by-year calibrations, respectively. Results for bias and R^2 of ET simulation are presented in Fig. S2 and S3. As the calibration and validation performances were similar and validation performance is more relevant to practical applications, we will only show and analyze the validation performance in calibration using BESS-STAIR ET.

3.2.2. Site-level calibration using BESS-STAIR ET product

Calibration using BESS-STAIR ET achieved similar performance with calibration using site ET (Fig. 4). With site forcing, RMSEs of ET simulation on average were reduced from 61 W/m^2 in default model simulation to 46 and 47 W/m^2 in multi-year calibration using site and BESS-STAIR ET, respectively, for maize, and from 66 W/m^2 in default model simulation to 53 and 53 W/m^2 for soybean. With regional forcing, RMSEs of ET simulation were reduced from 92 W/m^2 in default model

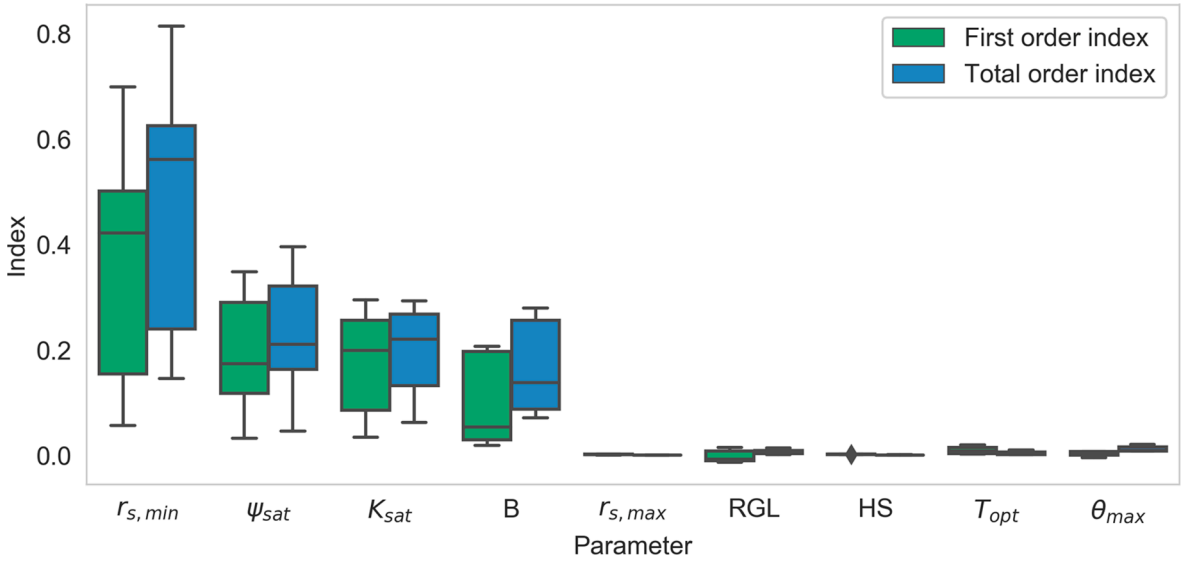

Fig. 2. Sobol sensitivity indices of 9 parameters across 8 sites.

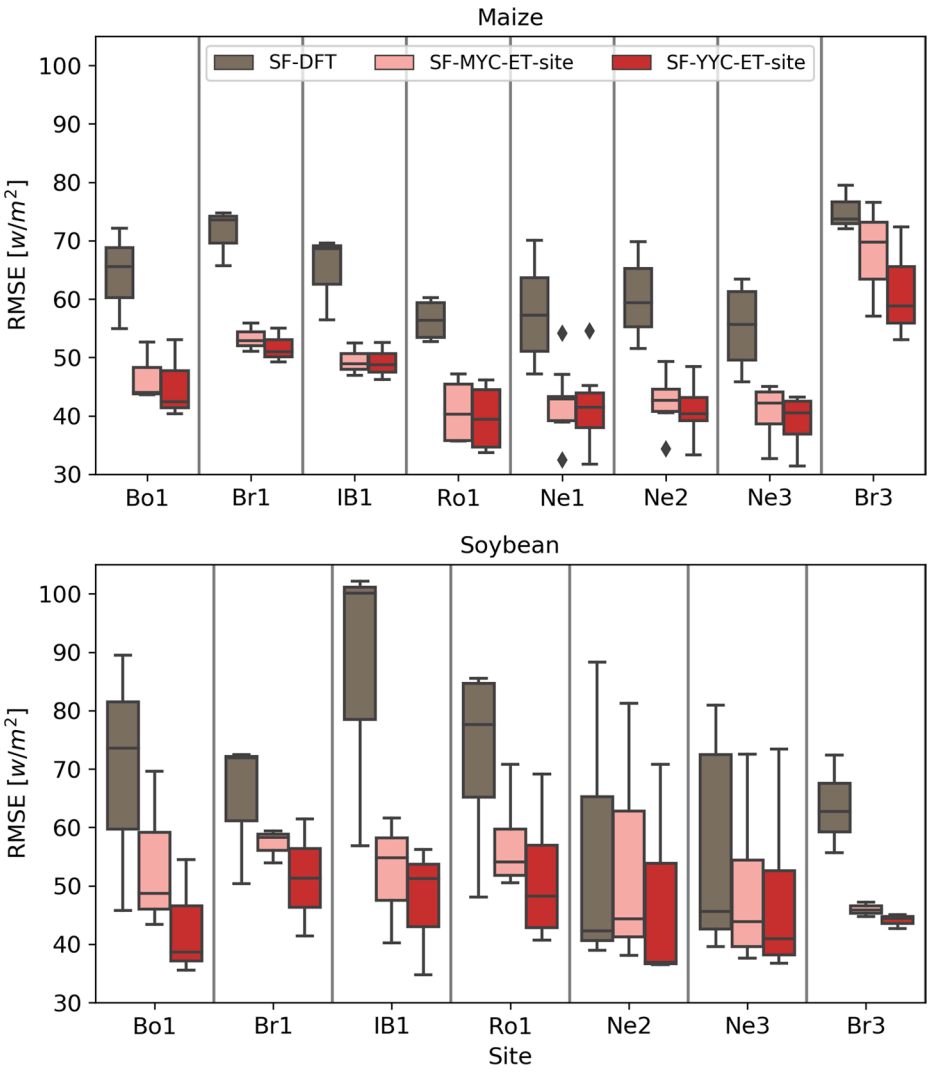

Fig. 3. The performance of ET simulation in different calibration strategies using flux tower measured ET. The three boxes for each site from left to right are: default model simulation with site forcing (brown; SF-DFT), multi-year calibration using site ET and with site forcing (pink; SF-MYC-ET-site; validation performance), year-by-year calibration using site ET and with site forcing (red; SF-YYC-ET-site; calibration performance). The results are shown separately for maize and soybean. Note that multi-year calibration gave validation performance, while year-by-year calibration gave calibration performance.

simulation to 63 and 64 W/m^2 in multi-year calibration using site and BESS-STAIR ET, respectively, for maize, and from 93 W/m^2 to 65 and 66 W/m^2 for soybean. The performance improvement was larger in calibration using BESS-STAIR ET with regional forcing (30% and 29% reduction of RMSEs of ET for maize and soybean, respectively) than with site forcing (23% and 20% reduction of RMSEs of ET for maize and soybean, respectively). The KGE performance is shown in Fig. S4. Comparison between simulated and observed ET time series show that calibration did not change much the short-term variations of ET; however, calibration was able to correct the bias of ET as well as improve its long-term dynamics (Fig. 5). In US-Ne3 2012, the simulated ET by default Noah-MP is higher in early growing season and lower in late growing season than observations, which was partially corrected by calibration.

At Ne1, Ne2 and Ne3, the performance of the simulated first layer SWC was respectable with little change after calibration (Table S1). With site forcing, there was barely any difference of ubRMSE before and after calibration; with regional forcing, the ubRMSE did improve from 0.43 to 0.36 or 0.37 after calibration using site ET or BESS-STAIR ET, respectively. However, the SWC performance improvement with regional forcing could be simply due to the transfer of uncertainties from forcing to calibrated parameters. Overall, the calibration had little effect on SWC simulations.

To further compare the difference of calibration using site ET and using BESS-STAIR ET, the calibrated parameters from the two calibrations are shown in Fig. 6. The three soil parameters were similar between the two calibrations; both the absolute values and the deviations from default values were mostly in line. For the vegetation parameter $r_{s,min}$, although the results showed some differences between calibration using site ET and BESS-STAIR ET across 8 sites, at Ne1, Ne2 and Ne3 sites where the forcing quality is high and there are relatively more data to get robust results, the calibrated $r_{s,min}$ were similar between the two calibrations for both maize and soybean. This indicated that calibrations using site ET and BESS-STAIR ET not only led to similar performance but also mostly similar parameters.

The performances of BESS-STAIR ET, simulated ET by calibrated Noah-MP models using BESS-STAIR and site ET, and the default Noah-MP model at Ne1, Ne2 and Ne3 were compared in Fig. 7. The performances were assessed using flux tower ET observations as the benchmark. All metrics were calculated using daily ET to match BESS-STAIR ET's temporal resolution, so the RMSE and R^2 are not comparable to those calculated using hourly ET in previous sections. We only did this comparison at Ne1, Ne2 and Ne3 because those sites have gap-free ET data so that they could be aggregated to daily scale. These results showed that (1) calibration using BESS-STAIR ET could match the performance of calibration directly using flux tower ET in terms of RMSE; (2) calibration improved the performance of model simulated ET mainly through the reduction of bias with less significant change in R^2. This comparison also revealed one major reason that BESS-STAIR ET could improve the model performance as much as site ET could is the low bias

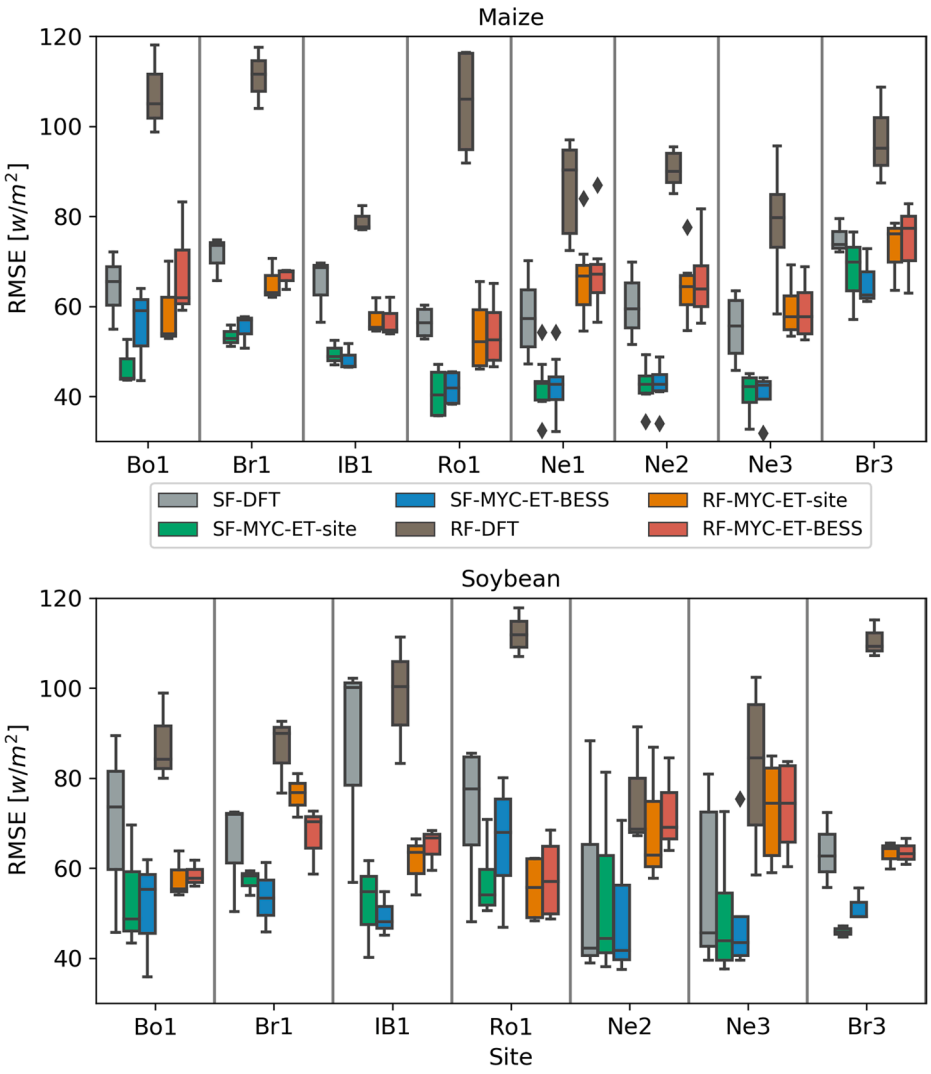

Fig. 4. Performance of multi-year calibration using BESS-STAIR ET compared to using site ET. The six boxes for each site from left to right are: default Noah-MP simulations with site forcing (SF-DFT), multi-year calibration using site ET with site forcing (SF-MYC-ET-site), multi-year calibration using BESS-STAIR ET with site forcing (SF-MYC-ET-BESS), default model simulation with regional forcing (RF-DFT), multi-year calibration using site ET with regional forcing (RF-MYC-ET-site), multi-year calibration using BESS-STAIR ET with regional forcing (RF-MYC-ET-BESS).

itself.

3.3. Regional calibration study

The performance of the simulated ET was significantly improved after the calibration across the region (Fig. 8 (a) and (c)). The average RMSE between simulated daily growing season (April 1 to October 31) ET and BESS-STAIR ET from 2014 to 2017 was reduced from 43 w/m2 to 25 w/m2. The performance improvement was primarily from the reduction of bias with moderate increase in R^2, which is in line with site-level studies. The water budget analysis shows that the calibration improved the regional water balance over the four years (Fig. 8 (b)). The model simulated total runoff can better match the observed data. This is because there was a positive bias of the simulated ET with default parameters, and calibration reduced the bias of ET, thus correcting the bias of simulated total runoff as well. We note that the runoff dynamics did not change much after calibration (Fig. S6), with little difference in terms of NSE coefficient. The comparison between the default model simulation, lumped and spatially explicit calibration shows that both lumped and spatially explicit calibration improved the performance of ET simulation and the distributed calibration provided better performance (Fig. S7). However, the distributed calibration did not show significant advantage over lumped calibration in terms of runoff performance (Fig. S8).

4. Discussions

4.1. Effectiveness of high-resolution model calibration using BESS-STAIR ET

The results of the calibration experiments addressed several questions regarding the effectiveness of high-resolution spatially explicit calibration using BESS-STAIR ET to improve the performance of Noah-MP LSM: (1) the calibration using BESS-STAIR ET can significantly improve Noah-MP's simulation of ET and mostly match the performance of calibrating directly using flux tower ET; (2) the field-scale ET performance can be significantly improved with the spatially explicit regional calibration; and (3) the regional calibration can also improve the long-term regional water budget simulation.

The site-level calibration experiments showed that calibrating the most sensitive parameters ($r_{s,min}$, K_{sat}, B, ψ_{sat}) using BESS-STAIR ET could significantly improve the ET performance of Noah-MP (Fig. 4). At the same time, the performance of the calibrated model using BESS-STAIR ET can match that of calibration directly using site ET in every aspect (Figs. 4, 6, 7). The minimal difference between multi-year and year-by-year calibration indicated similar performance between calibration and validation, suggesting that the calibrated parameters can be well applied to other years when no ET observations (either from site or satellite) are available.

The regional calibration study on the Spoon River watershed

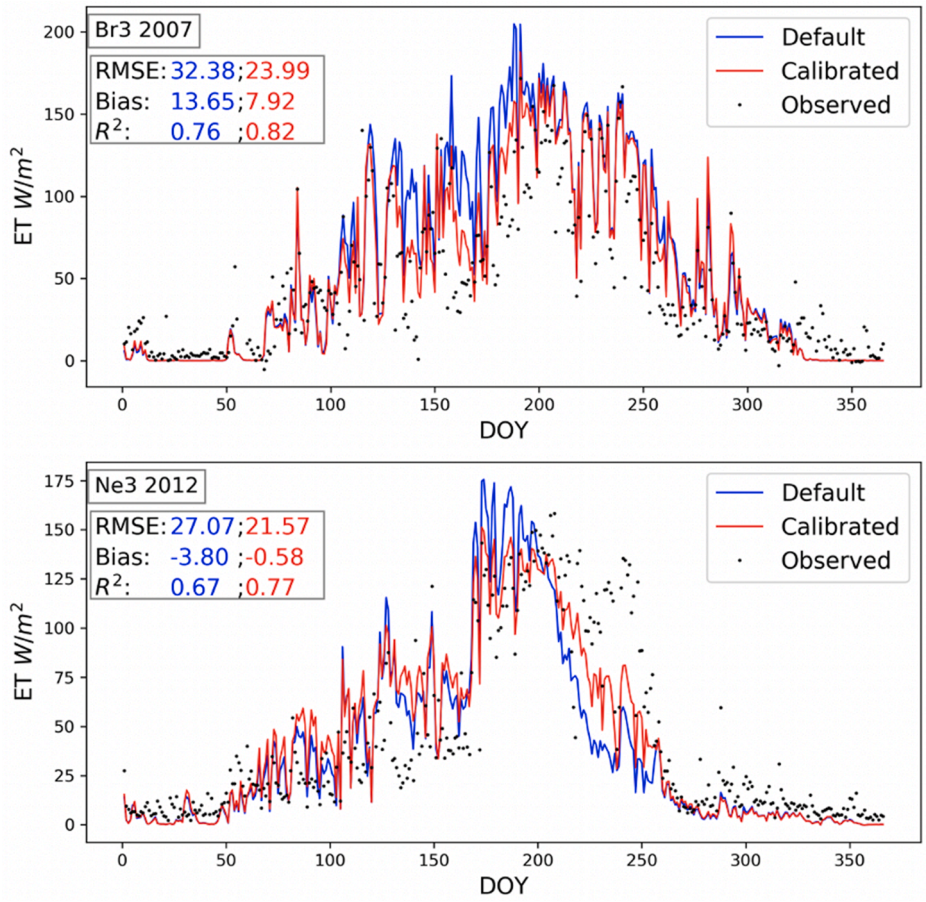

Fig. 5. Time series of simulated daily ET before and after calibration using BESS-STAIR ET (multi-year calibration, which reflects validation performance). Top: simulation in Br3, 2007; bottom: simulation in Ne3, 2012.

demonstrated that field-scale ET simulation can be improved through spatially explicit calibration using high-resolution data, and better performance can be achieved than traditional lumped calibration only using streamflow. ET, like many surface fluxes, has large spatial heterogeneities, especially in agricultural fields. In land surface models, the heterogeneities are primarily controlled by heterogeneous vegetation and soil parameters. The high-resolution spatially explicit calibration could be an effective way to capture and constrain the spatial heterogeneities of those parameters and improve land surface modeling at fine scales. Improved long-term regional water budget simulation was also observed after the calibration, which further validated the effectiveness of the regional calibration experiment.

4.2. Implications for hyper-resolution modeling and model-data integration

The regional calibration study demonstrated the effectiveness of high-resolution calibration using BESS-STAIR ET in improving the simulation of Noah-MP and provided a concrete example of applying the calibration framework to agricultural fields. Theoretically, calibration can be done at the same 30 m resolution as BESS-STAIR ET itself; however, for agricultural applications, this regional calibration study was done at individual field level using extracted field boundaries to reduce computational cost. Field boundary extraction is an important technique that can help better organize the heterogeneity in land surface models over agricultural areas (Yan and Roy, 2016, 2014). In addition, USDA CDL data was used to distinguish different crop types at field scale, which is largely absent in previous hydrological model simulation and calibration studies. Beyond the current regional study, other spatially distributed hydrological/land surface variables can be potentially combined with ET for multi-objective calibration to better constrain the model. Furthermore, a dynamic crop model has been incorporated into Noah-MP recently (Liu et al., 2016), so future work could investigate if the simulation of crop photosynthesis and yield (Guan et al., 2017) could be improved with the constraint of BESS-STAIR ET.

Compared with previous studies, the advantage of this study lies first and foremost in the high resolution of the satellite-based data. Previous spatially explicit calibration studies were done at coarser resolutions like 36 km with SMOS soil moisture data (Shellito et al., 2016) or even sub-basin scale with MODIS ET (Rajib et al., 2018). The 30-m resolution BESS-STAIR ET product is currently available primarily across the U.S. Corn Belt, and can be applied to constrain models in those agricultural landscapes. Being able to resolve individual fields at this resolution, the framework could be especially useful for precision agricultural applications. Another important advantage of this study is the efficacy of the remote sensing data in calibration. As we have demonstrated, the low bias feature of BESS-STAIR ET (Fig. 7) makes it a viable substitute for the ground truth data (e.g. eddy covariance flux data) in calibration at large scales. This is in contrast with previous studies which either did not compare with calibration using ground truth data or found the satellite-based data like SMOS soil moisture need to be bias corrected for it to be useful in improving model simulations through calibration (Shellito et al., 2016).

That the efficacy of BESS-STAIR ET in calibration lied to a large extent in its low bias could be due to two possible reasons. Firstly, three of the calibrated parameters were soil hydraulic properties. Soil moisture dynamics in the root zone is a relatively slowly varying process

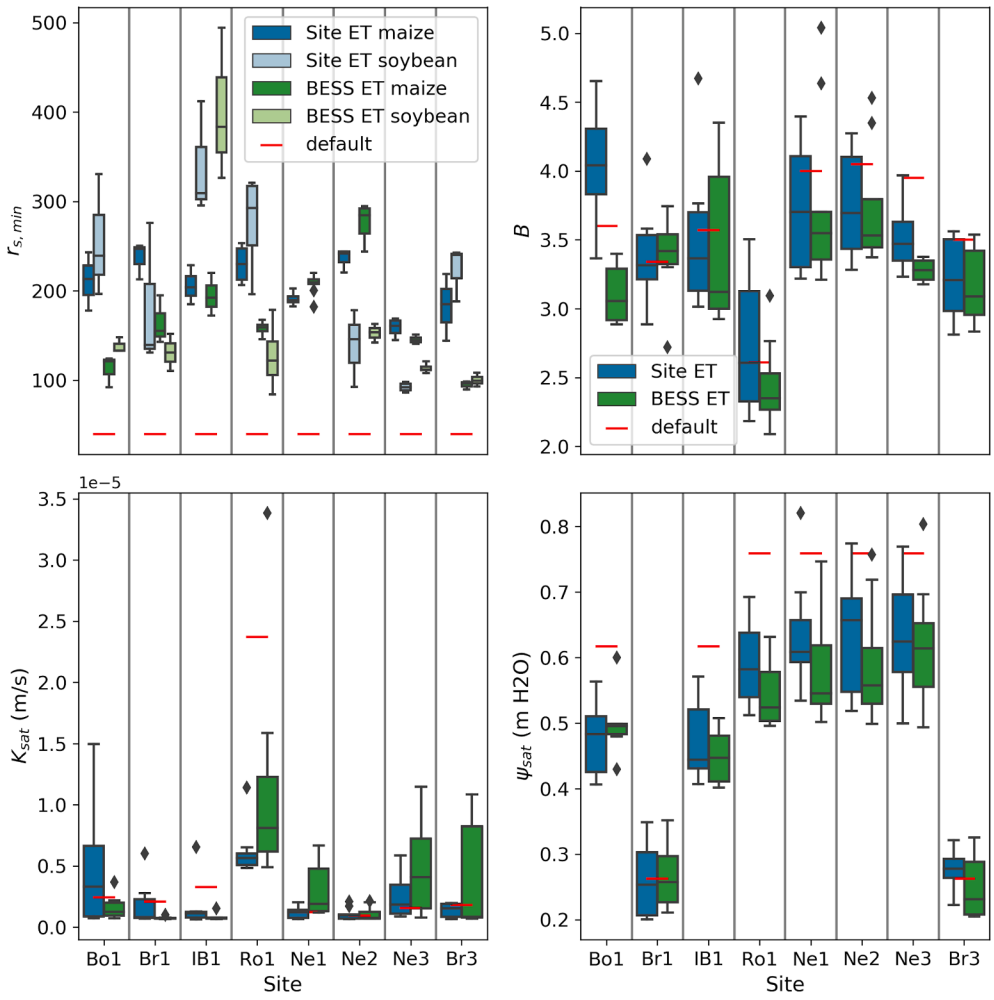

Fig. 6. Calibrated parameter values from multi-year calibration experiments using both site ET and BESS-STAIR ET.

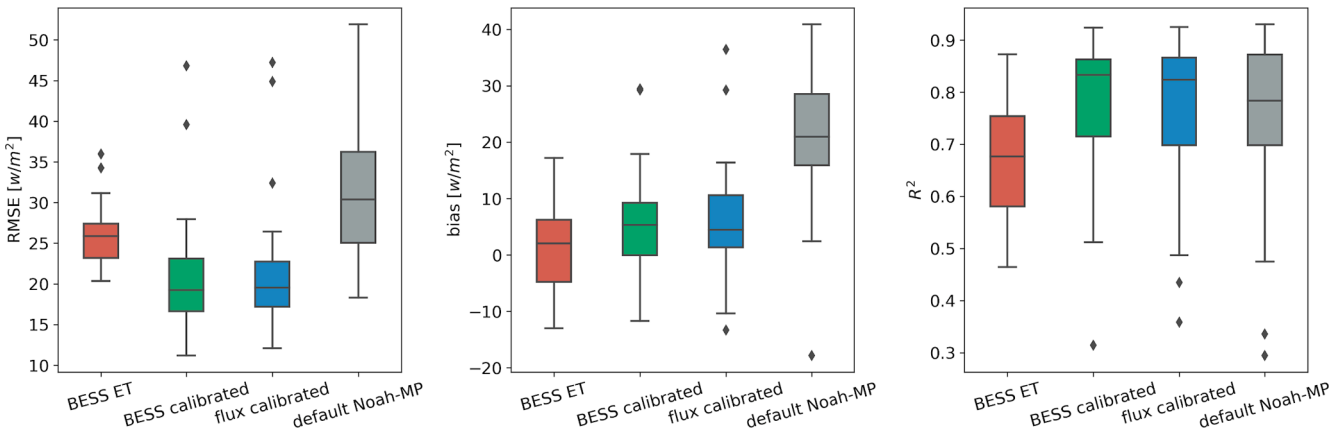

Fig. 7. Performance inter-comparison of daily ET from BESS-STAIR, calibrated and default Noah-MP model simulations with site forcing at US-Ne1, US-Ne2 and US-Ne3. The four boxes from left to right represent: BESS-STAIR ET, simulated ET by calibrated Noah-MP models using BESS-STAIR and site ET, and default Noah-MP model.

compared to other meteorological processes driving ET; thus, the soil moisture control is likely to play a bigger role in seasonal variations than diurnal or day-to-day variations of ET. As the model performance was calibrated and evaluated at either hourly or daily scale in our study, the correlation performance was dominated by diurnal or day-to-day variations. Therefore, calibration of soil parameters leads to more improvements to the RMSE or bias performance, instead of correlation performance. Another possible reason is that the calibrated vegetation parameter $r_{s,min}$ only changed the magnitude of stomatal resistance without changing its dynamics (Eq. (3)), so calibrating $r_{s,min}$ would barely respond to the correlation performance either. Accordingly, we can reasonably infer that as long as the satellite-based data performs sufficiently good at longer time scales and can bring the model bias down to the same level as ground truth data does through calibration

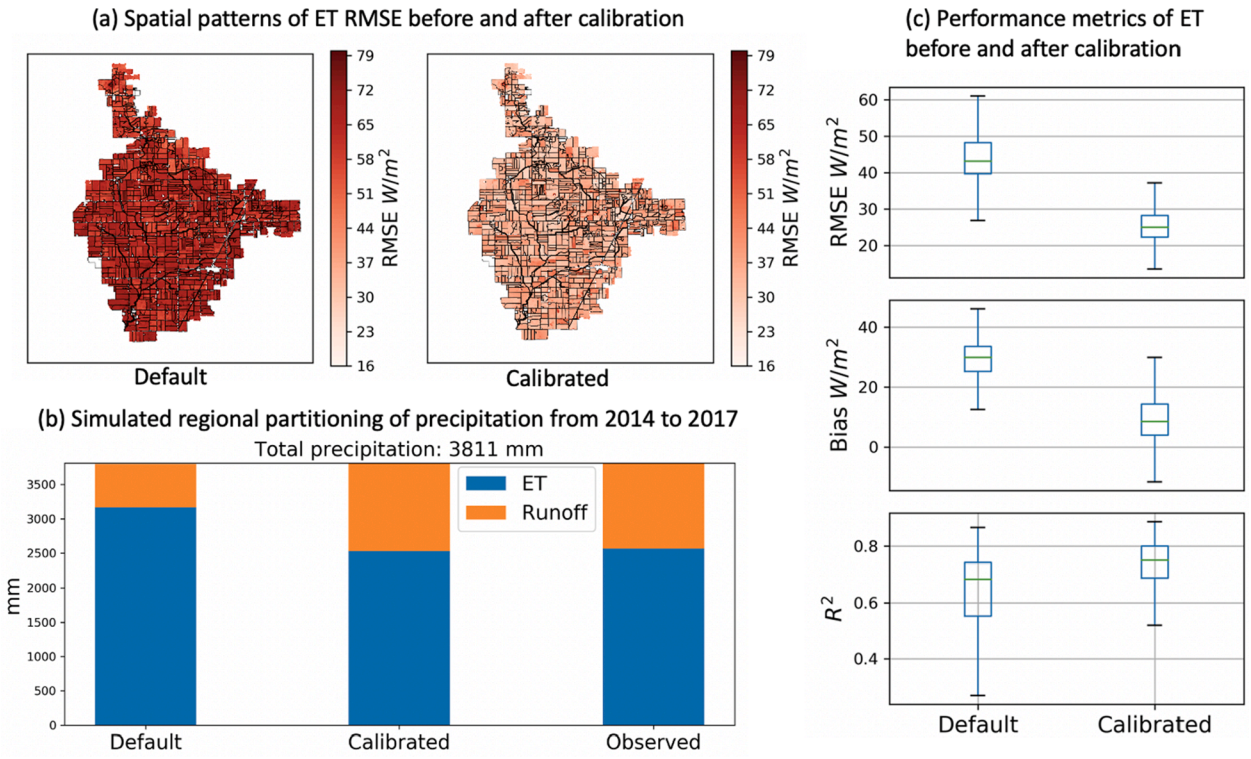

Fig. 8. Performance of regional simulation from 2014 to 2017 before and after calibration. Performance metrics of ET simulation were calculated against BESS-STAIR ET. In water budget analysis, changes of soil water and groundwater were negligible so they are not reflected in the chart, and the observed ET was calculated by subtracting observed runoff from precipitation.

without hurting the correlation performance, the satellite-based data can match the performance of ground truth data in calibration. BESS-STAIR ET satisfies the conditions because of its low bias and excellent seasonal performance (Jiang et al., 2019).

Encouragingly, large sources of uncertainties in ET simulation in land surface models are from stomatal and soil moisture controls on ET (Egea et al., 2011; Keenan et al., 2010), which are partially reflected in the 4 parameters in this study. As demonstrated in the calibration experiments, the uncertainties in these parameters could be well constrained by using satellite-based BESS-STAIR ET estimation even if the satellite-based ET estimation could not fully match the temporal dynamics of flux tower observed ET at short time scales. This provides an insight into model-data integration that combining the strengths of model simulations and satellite-based data at different time scales could potentially provide better estimations. For instance, as shown in this study, when driven by good quality surface forcing, default Noah-MP simulated ET could already achieve good correlation with flux tower measurements, and calibration using BESS-STAIR ET further improved the performance by reducing the bias (Fig. 7). As a result, the calibrated model could have a superior performance than both BESS-STAIR ET and default Noah-MP in terms of RMSE performance (Fig. 7).

4.3. Limitations and potential improvement

Several ways exist to potentially improve the current study. We used the Jarvis type stomatal resistance scheme for simplicity, but it is insufficient to simulate many important processes, such as canopy-scale photosynthesis (we turned it off in this study). Ball-Berry type (Ball et al., 1987) or even more advanced stomatal resistance schemes (Manzoni et al., 2011; Medlyn et al., 2011) could be more suitable for simulating canopy carbon uptake, which would enable using ET to constrain parameters related with canopy carbon assimilation. Another missing part in the Noah-MP model was the artificial tile drainage prevalent in some parts of the U.S. Midwest, which can modify hydrological and biogeochemical processes significantly across the agricultural landscape (Gentry et al., 2007; Green et al., 2006). Thus, the calibrated soil hydraulic parameters in this study were essentially "effective" parameters, not fully representing the soil physical properties. Although our model calibration scheme could lead to improved ET simulation, we admit that the processes related with tile drainage need to be explicitly represented in the model for realistic simulation of the agricultural ecosystems in the U.S. Midwest (Boles et al., 2015; Guo et al., 2018; Li et al., 2010). The relationship between ET and soil moisture is another source of model uncertainty that could compromise the results of the calibration study. Noah-MP uses an empirical linear soil water stress function to attenuate ET when soil moisture is low. Studies have shown that this type of approaches tend to overestimate ET sensitivity to insufficient soil moisture. As shown in Table S1, soil moisture dynamics has little improvement after calibration. If the soil moisture–ET relationship is improved, the high-resolution ET data could be potentially better utilized to improve the dynamics of other water cycle components including soil moisture. Therefore, despite the increasing availability of high-resolution and high-accuracy remote sensing data, land surface models still need to be continuously improved. We also note that the calibration conducted in this study hardly improved model dynamics, as demonstrated by the unchanged R^2 performance even after calibration using flux tower measured ET, nor did it improve the simulation of SWC and other energy balance terms except ET. In the regional calibration experiments, although the long-term water budget simulation was improved under the constraint of the high-resolution ET, it is still difficult to improve runoff dynamics through the distributed calibration. The NSE coefficient of simulated streamflow evaluated with in-situ measurements is relatively low regardless of calibration. We also note that those fields that have higher RMSE before calibration still have relatively higher RMSE after calibration, which means even the spatially explicit calibration cannot fully resolve the field scale dynamics. Aside from model physics improvement, more advanced model-data integration techniques, such as multi-

objective calibration (Engeland et al., 2006; Yapo et al., 1998) and constraining emergent relationships (Hall et al., 2019; Peng et al., 2020) could be beneficial for reducing uncertainties of land surface models. In sum, field scale hydrological modeling remains a challenge and significant work needs to be done to resolve field scale dynamics of water and energy cycles.

5. Conclusion

In this study, we investigated the effectiveness of high-resolution spatially explicit parameter calibration using BESS-STAIR ET and the Noah-MP land surface model. Site-level calibration experiments at eight flux tower sites in the U.S. Corn Belt and a regional study in the Soon River watershed were conducted. We demonstrated that the high-resolution calibration using BESS-STAIR ET can improve the ET simulation of Noah-MP and found that calibration using BESS-STAIR ET could match the results of calibration directly using flux tower measurements. The performance of field-scale ET and long-term regional water budget simulation was also observed in regional calibration experiments. The high-resolution distributed calibration offered superior performance than traditional lumped calibration using streamflow only. Further analysis revealed that BESS-STAIR ET's low bias and excellent long-term correlation performance could be the main contributors to its efficacy in constraining the model, and the low bias performance of satellite-based data and good correlation performance of model simulation can be potentially combined through model-data integration. Overall, our study provided a concrete example of improving field-scale land surface modeling through spatially explicit calibration using high-resolution satellite data in agricultural landscapes.

CRediT authorship contribution statement

Yi Yang: Formal analysis, Investigation, Writing - original draft. **Kaiyu Guan:** Conceptualization, Methodology, Writing - review & editing. **Bin Peng:** Conceptualization, Methodology, Writing - review & editing. **Ming Pan:** Methodology, Writing - review & editing. **Chongya Jiang:** Methodology, Writing - review & editing. **Trenton E. Franz:** Methodology, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

We acknowledge the support from NSF CAREER award (Award Abstract #1847334) managed through the NSF Environmental Sustainability Program and USDA/NSF Cyber-Physical-System Program. We also acknowledge the support from the annual small research grant for graduate students from USGS Illinois Water Resources Center. We thank Murugesu Sivapalan for helpful comments on this study. We acknowledge the following AmeriFlux sites for their data records: US-Ne1, US-Ne1, US-Ne3, US- Bo1, US-IB1, US-Br1, and US-Ro1, US-Br3. In addition, funding for AmeriFlux data resources and core site data was provided by the U.S. Department of Energy's Office of Science.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2020.125730.

References

Anderson, M.C., Kustas, W.P., Norman, J.M., Hain, C.R., Mecikalski, J.R., Schultz, L., González-Dugo, M.P., Cammalleri, C., D'Urso, G., Pimstein, A., Gao, F., 2011. Mapping daily evapotranspiration at field to continental scales using geostationary and polar orbiting satellite imagery. Hydrol. Earth Syst. Sci. 15, 223–239. https://doi.org/10.5194/hess-15-223-2011.

Austin, A.T., Yahdjian, L., Stark, J.M., Belnap, J., Porporato, A., Norton, U., Ravetta, D. A., Schaeffer, S.M., 2004. Water pulses and biogeochemical cycles in arid and semiarid ecosystems. Oecologia 141, 221–235. https://doi.org/10.1007/s00442-004-1519-1.

Ball, J.T., Woodrow, I.E., Berry, J.A., 1987. A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions. In: Progress in Photosynthesis Research. Springer, Netherlands, Dordrecht, pp. 221–224. https://doi.org/10.1007/978-94-017-0519-6_48.

Beck, H.E., Pan, M., Roy, T., Weedon, G.P., Pappenberger, F., Van Dijk, A.I.J.M., Huffman, G.J., Adler, R.F., Wood, E.F., 2019. Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. Hydrol. Earth Syst. Sci. 23, 207–224. https://doi.org/10.5194/hess-23-207-2019.

Beck, H.E., van Dijk, A.I.J.M., de Roo, A., Miralles, D.G., McVicar, T.R., Schellekens, J., Bruijnzeel, L.A., 2016. Global-scale regionalization of hydrologic model parameters. Water Resour. Res. 52, 3599–3622. https://doi.org/10.1002/2015WR018247.

Beven, K., 2006. A manifesto for the equifinality thesis. J. Hydrol. 320, 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007.

Boles, C.M.W., Frankenberger, J.R., Moriasi, D.N., 2015. Tile drainage simulation in SWAT2012: parameterization and evaluation in an indiana watershed. Trans. ASABE 58, 1201–1213. https://doi.org/10.13031/trans.58.10589.

Chaney, N.W., Herman, J.D., Ek, M.B., Wood, E.F., 2016. Deriving global parameter estimates for the Noah land surface model using FLUXNET and machine learning. J. Geophys. Res. 121, 13218–13235. https://doi.org/10.1002/2016JD024821.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., 2013. Carbon and other biogeochemical cycles, in: Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 465–570. https://doi.org/10.1017/CBO9781107415324.015.

Crow, W.T., Wood, E.F., Pan, M., 2003. Multiobjective calibration of land surface model evapotranspiration predictions using streamflow observations and spaceborne surface radiometric temperature retrievals. J. Geophys. Res. D Atmos. 108, 1–12. https://doi.org/10.1029/2002jd003292.

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour. Res. 28, 1015–1031. https://doi.org/10.1029/91WR02985.

Duan, Q.Y., Gupta, V.K., Sorooshian, S., 1993. Shuffled complex evolution approach for effective and efficient global minimization. J. Optim. Theory Appl. 76, 501–521. https://doi.org/10.1007/BF00939380.

Egea, G., Verhoef, A., Vidale, P.L., 2011. Towards an improved and more flexible representation of water stress in coupled photosynthesis-stomatal conductance models. Agric. For. Meteorol. 151, 1370–1384. https://doi.org/10.1016/j.agrformet.2011.05.019.

Engeland, K., Braud, I., Gottschalk, L., Leblois, E., 2006. Multi-objective regional modelling. J. Hydrol. 327, 339–351. https://doi.org/10.1016/j.jhydrol.2005.11.022.

Gentry, L.E., David, M.B., Royer, T.V., Mitchell, C.A., Starks, K.M., 2007. Phosphorus transport pathways to streams in tile-drained agricultural watersheds. J. Environ. Qual. 36, 408–415. https://doi.org/10.2134/jeq2006.0098.

Green, C.H., Tomer, M.D., Di Luzio, M., Arnold, J.G., 2006. Hydrologic evaluation of the soil and water assessment tool for a large tile-drained watershed in Iowa. Trans. ASABE 49, 413–422.

Guan, K., Wu, J., Kimball, J.S., Anderson, M.C., Frolking, S., Li, B., Hain, C.R., Lobell, D. B., 2017. The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. Remote Sens. Environ. 199, 333–349. https://doi.org/10.1016/j.rse.2017.06.043.

Guo, T., Gitau, M., Merwade, V., Arnold, J., Srinivasan, R., Hirschi, M., Engel, B., 2018. Comparison of performance of tile drainage routines in SWAT 2009 and 2012 in an extensively tile-drained watershed in the Midwest. Hydrol. Earth Syst. Sci. 22, 89–110. https://doi.org/10.5194/hess-22-89-2018.

Gutmann, E.D., Small, E.E., 2010. A method for the determination of the hydraulic properties of soil from MODIS surface temperature for use in land-surface models. Water Resour. Res. 46, 1–16. https://doi.org/10.1029/2009WR008203.

Hall, A., Cox, P., Huntingford, C., Klein, S., 2019. Progressing emergent constraints on future climate change. Nat. Clim. Change. 9, 269–278. https://doi.org/10.1038/s41558-019-0436-6.

Houser, P.R., Shuttleworth, W.J., Famiglietti, J.S., Gupta, H.V., Syed, K.H., Goodrich, D. C., 1998. Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. Water Resour. Res. 34, 3405–3420. https://doi.org/10.1029/1998WR900001.

Immerzeel, W.W., Droogers, P., 2008. Calibration of a distributed hydrological model based on satellite evapotranspiration. J. Hydrol. 349, 411–424. https://doi.org/10.1016/j.jhydrol.2007.11.017.

Jackson, T.J., O'Neill, P.E., Swift, C.T., 1997. Passive microwave observation of diurnal surface soil moisture. IEEE Trans. Geosci. Remote Sens. 35, 1210–1222. https://doi.org/10.1109/36.628788.

Jarvis, P.G., 1976. The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. Philos. Trans. R. Soc. London. B Biol. Sci. 273, 593–610. https://doi.org/10.1098/rstb.1976.0035.

Jiang, C., Guan, K., Pan, M., Ryu, Y., Peng, B., Wang, S., 2019. BESS-STAIR: a framework to estimate daily, 30-meter, and allweather crop evapotranspiration using multi-source satellite data for the U.S. Corn Belt. Hydrol. Earth Syst. Sci. Discuss. 1–36 https://doi.org/10.5194/hess-2019-376.

Jiang, C., Ryu, Y., 2016. Multi-scale evaluation of global gross primary productivity and evapotranspiration products derived from Breathing Earth System Simulator (BESS). Remote Sens. Environ. 186, 528–547. https://doi.org/10.1016/j.rse.2016.08.030.

Karthikeyan, L., Chawla, I., Mishra, A.K., 2020. A review of remote sensing applications in agriculture for food security: crop growth and yield, irrigation, and crop losses. J. Hydrol. 586, 124905 https://doi.org/10.1016/j.jhydrol.2020.124905.

Keenan, T., Sabate, S., Gracia, C., 2010. Soil water stress and coupled photosynthesis-conductance models: bridging the gap between conflicting reports on the relative roles of stomatal, mesophyll conductance and biochemical limitations to photosynthesis. Agric. For. Meteorol. 150, 443–453. https://doi.org/10.1016/j.agrformet.2010.01.008.

Kerr, Y.H., Waldteufel, P., Wigneron, J.P., Martinuzzi, J.M., Font, J., Berger, M., 2001. Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission. IEEE Trans. Geosci. Remote Sens. 39, 1729–1735. https://doi.org/10.1109/36.942551.

Ledieu, J., De Ridder, P., De Clerck, P., Dautrebande, S., 1986. A method of measuring soil moisture by time-domain reflectometry. J. Hydrol. 88, 319–328. https://doi.org/10.1016/0022-1694(86)90097-1.

Li, H., Sivapalan, M., Tian, F., Liu, D., 2010. Water and nutrient balances in a large tile-drained agricultural catchment: a distributed modeling study. Hydrol. Earth Syst. Sci. 14, 2259–2275. https://doi.org/10.5194/hess-14-2259-2010.

Li, Z.L., Tang, R., Wan, Z., Bi, Y., Zhou, C., Tang, B., Yan, G., Zhang, X., 2009. A review of current methodologies for regional Evapotranspiration estimation from remotely sensed data. Sensors 9, 3801–3853. https://doi.org/10.3390/s90503801.

Liou, Y.A., Kar, S.K., 2014. Evapotranspiration estimation with remote sensing and various surface energy balance algorithms-a review. Energies 7, 2821–2849. https://doi.org/10.3390/en7052821.

Liu, X., Chen, F., Barlage, M., Zhou, G., Niyogi, D., 2016. Noah-MP-Crop: introducing dynamic crop growth in the Noah-MP land surface model. J. Geophys. Res. 121, 13953–13972. https://doi.org/10.1002/2016JD025597.

Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based crop yield mapper. Remote Sens. Environ. 164, 324–333. https://doi.org/10.1016/j.rse.2015.04.021.

Luo, Y., Guan, K., Peng, J., 2018. STAIR: A generic and fully-automated method to fuse multiple sources of optical satellite data to generate a high-resolution, daily and cloud-/gap-free surface reflectance product. Remote Sens. Environ. 214, 87–99. https://doi.org/10.1016/j.rse.2018.04.042.

Manzoni, S., Vico, G., Katul, G., Fay, P.A., Polley, W., Palmroth, S., Porporato, A., 2011. Optimizing stomatal conductance for maximum carbon gain under water stress: a meta-analysis across plant functional types and climates. Funct. Ecol. 25, 456–467. https://doi.org/10.1111/j.1365-2435.2010.01822.x.

Medlyn, B.E., Duursma, R.A., Eamus, D., Ellsworth, D.S., Prentice, I.C., Barton, C.V.M., Crous, K.Y., De Angelis, P., Freeman, M., Wingate, L., 2011. Reconciling the optimal and empirical approaches to modelling stomatal conductance. Glob. Chang. Biol. 17, 2134–2144. https://doi.org/10.1111/j.1365-2486.2010.02375.x.

Niraula, R., Norman, L.M., Meixner, T., Callegary, J.B., 2012. Multi-gauge calibration for modeling the semi-arid santa cruz watershed in Arizona-Mexico border area using SWAT. Air, Soil Water Res. 5, 41–57. https://doi.org/10.4137/ASWR.S9410.

Niu, G.Y., Yang, Z.L., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., Xia, Y., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. J. Geophys. Res. Atmos. 116, 1–19. https://doi.org/10.1029/2010JD015139.

Pan, M., Cai, X., Chaney, N.W., Entekhabi, D., Wood, E.F., 2016. An initial assessment of SMAP soil moisture retrievals using high-resolution model simulations and in situ observations. Geophys. Res. Lett. 43, 9662–9668. https://doi.org/10.1002/2016GL069964.

Peng, B., Guan, K., Tang, J., Ainsworth, E.A., Asseng, S., Bernacchi, C.J., Cooper, M., Delucia, E.H., Elliott, J.W., Ewert, F., Grant, R.F., Gustafson, D.I., Hammer, G.L., Jin, Z., Jones, J.W., Kimm, H., Lawrence, D.M., Li, Y., Lombardozzi, D.L., Marshall-Colon, A., Messina, C.D., Ort, D.R., Schnable, J.C., Vallejos, C.E., Wu, A., Yin, X., Zhou, W., 2020. Towards a multiscale crop modelling framework for climate change adaptation assessment. Nat. Plants 6, 338–348. https://doi.org/10.1038/s41477-020-0625-3.

Peng, B., Zhao, T., Shi, J., Lu, H., Mialon, A., Kerr, Y.H., Liang, X., Guan, K., 2017. Reappraisal of the roughness effect parameterization schemes for L-band radiometry over bare soil. Remote Sens. Environ. 199, 63–77. https://doi.org/10.1016/j.rse.2017.07.006.

Rajib, A., Evenson, G.R., Golden, H.E., Lane, C.R., 2018. Hydrologic model predictability improves with spatially explicit calibration using remotely sensed evapotranspiration and biophysical parameters. J. Hydrol. 567, 668–683. https://doi.org/10.1016/j.jhydrol.2018.10.024.

Ryu, Y., Baldocchi, D.D., Kobayashi, H., Van Ingen, C., Li, J., Black, T.A., Beringer, J., Van Gorsel, E., Knohl, A., Law, B.E., Roupsard, O., 2011. Integration of MODIS land and atmosphere products with a coupled-process model to estimate gross primary productivity and evapotranspiration from 1 km to global scales. Global Biogeochem. Cycles 25, 1–24. https://doi.org/10.1029/2011GB004053.

Ryu, Y., Jiang, C., Kobayashi, H., Detto, M., 2018. MODIS-derived global land products of shortwave radiation and diffuse and total photosynthetically active radiation at 5 km resolution from 2000. Remote Sens. Environ. 204, 812–825. https://doi.org/10.1016/j.rse.2017.09.021.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun. 181, 259–270. https://doi.org/10.1016/j.cpc.2009.09.018.

Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resour. Res. 46, 1–25. https://doi.org/10.1029/2008WR007327.

Schaake, J.C., Koren, V.I., Duan, Q.Y., Mitchell, K., Chen, F., 1996. Simple water balance model for estimating runoff at different spatial and temporal scales. J. Geophys. Res. Atmos. 101, 7461–7475. https://doi.org/10.1029/95JD02892.

Shellito, P.J., Small, E.E., Cosh, M.H., 2016. Calibration of Noah soil hydraulic property parameters using surface soil moisture from SMOS and basinwide in situ observations. J. Hydrometeorol. 17, 2275–2292. https://doi.org/10.1175/JHM-D-15-0153.1.

Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. Comput. Simul. 55, 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6.

Sorooshian, S., Duan, Q., Gupta, V.K., 1993. Calibration of rainfall-runoff models: Application of global optimization to the Sacramento Soil Moisture Accounting Model. Water Resour. Res. 29, 1185–1194. https://doi.org/10.1029/92WR02617.

Su, Z., 2002. The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. Hydrol. Earth Syst. Sci. 6, 85–99. https://doi.org/10.5194/hess-6-85-2002.

Sutanudjaja, E.H., Van Beek, L.P.H., De Jong, S.M., Van Geer, F.C., Bierkens, M.F.P., 2014. Calibrating a large-extent high-resolution coupled groundwater-land surface model using soil moisture and discharge data. Water Resour. Res. 50, 687–705. https://doi.org/10.1002/2013WR013807.

Velpuri, N.M., Senay, G.B., Singh, R.K., Bohms, S., Verdin, J.P., 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: using point and gridded FLUXNET and water balance ET. Remote Sens. Environ. 139, 35–49. https://doi.org/10.1016/j.rse.2013.07.013.

Vuichard, N., Papale, D., 2015. Filling the gaps in meteorological continuous data measured at FLUXNET sites with ERA-Interim reanalysis. Earth Syst. Sci. Data 7, 157–171. https://doi.org/10.5194/essd-7-157-2015.

Wanders, N., Bierkens, M.F.P., de Jong, S.M., de Roo, A., Karssenberg, D., 2014. The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological modelss. Water Resour. Res. 50, 6874–6891. https://doi.org/10.1002/2013WR014639.

Wood, E.F., Coauthors, 2011. Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. Water Resour. Res 47.

Xia, Y., Pitman, A.J., Gupta, H.V., Leplastrier, M., Henderson-Sellers, A., Bastidas, L.A., 2002. Calibrating a land surface model of varying complexity using multicriteria methods and the Cabauw dataset. J. Hydrometeorol. 3, 181–194.

Yan, L., Roy, D.P., 2016. Conterminous United States crop field size quantification from multi-temporal Landsat data. Remote Sens. Environ. 172, 67–86. https://doi.org/10.1016/j.rse.2015.10.034.

Yan, L., Roy, D.P., 2014. Automated crop field extraction from multi-temporal Web Enabled Landsat Data. Remote Sens. Environ. 144, 42–64. https://doi.org/10.1016/j.rse.2014.01.006.

Yang, Y., Pan, M., Beck, H.E., Fisher, C.K., Beighley, R.E., Kao, S.C., Hong, Y., Wood, E.F., 2019. In quest of calibration density and consistency in hydrologic modeling: distributed parameter calibration against streamflow characteristics. Water Resour. Res. 55, 7784–7803. https://doi.org/10.1029/2018wr024178.

Yang, Z.L., Niu, G.Y., Mitchell, K.E., Chen, F., Ek, M.B., Barlage, M., Longuevergne, L., Manning, K., Niyogi, D., Tewari, M., Xia, Y., 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins. J. Geophys. Res. Atmos. 116, 1–16. https://doi.org/10.1029/2010JD015140.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. J. Hydrol. 204, 83–97. https://doi.org/10.1016/S0022-1694(97)00107-8.