https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

Evaluation of water, snow, energy, and carbon dynamics in kilometer-scale E3SM Land Model (ELMv2) across the Contiguous United States

Gautam Bisht[1], Lingcheng Li[1], Donghui Xu[1], Dalei Hao[1], Hoang Tran[1], and L. Ruby Leung[1]
[1]Atmospheric, Climate, & Earth Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA
Correspondence: Gautam Bisht (gautam.bisht@pnnl.gov)

Abstract. Kilometer-scale (k-scale) land surface modeling is increasingly recognized as essential for accurately simulating terrestrial water, energy, and carbon cycles. In this study, we evaluate and improve the performance of the E3SM Land Model version 2 (ELMv2), the land component of E3SM version 2, configured at 1 km resolution across the Contiguous United States (CONUS). Offline ELMv2 simulations incorporating a newly developed 1 km surface dataset and improved parameterizations for sub-grid topographic effects on solar radiation and pluvial inundation were performed for 2001–2019, driven by NLDAS2 atmospheric forcing. Two uncertain model parameters, the subsurface drainage parameter (f_{drain}) and the pluvial flood dynamics parameter (f_c), were sequentially calibrated against benchmark water table depth and surface water dynamic datasets, leading to substantial improvements for both variables relative to the default configuration. Additional ten independent datasets were used to evaluate ELM-simulated water, snow, near-surface energy, and carbon cycle dynamics. The calibrated ELM captured the broad spatial and seasonal patterns of soil moisture, runoff, snow water equivalent (SWE), latent heat flux, land surface temperature, and gross primary productivity (GPP) reasonably well, though systematic biases remained, including a wet bias in soil moisture over the Midwest and Northeast, underestimated and muted-amplitude runoff with a shifted seasonal peak, underprediction of SWE across the mountainous West, an overestimation of summertime latent heat flux over the south-central CONUS, and an overprediction of GPP over the central and eastern domain. Relative to the simulation with default f_{drain} and f_c values, calibration had minimal impact on model performance against the ten independent benchmark datasets. Upscaling the 1 km simulation to 12 km resulted in substantial decrease in the information loss metric (34–44% on average) in net radiation, latent heat flux, sensible heat flux, and near-surface soil moisture, highlighting the value of k-scale land surface modeling for resolving fine-scale terrestrial processes. These results demonstrate the value of k-scale ELM simulations for resolving fine-scale spatial heterogeneity in water, energy, and carbon fluxes, while also highlighting the continued need for improved model physics, calibration strategies, and higher-fidelity input datasets to further reduce systematic biases as ELM moves toward global k-scale application.

1 Introduction

The need for a kilometer-scale (k-scale) resolution in land surface models (LSMs) to simulate the terrestrial water, energy, and carbon cycles is well recognized. More than a decade ago, Wood et al. (2011) articulated the grand challenge of using a hyper-

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

resolution of 1 km global land surface modeling because coarse-resolution LSMs cannot resolve the fine-scale heterogeneity in topography, soils, vegetation, and hydrology that governs the surface–subsurface and land–atmosphere exchanges of water and energy. Sub-grid processes, such as topographic controls on incoming solar radiation, snow accumulation and melt, pluvial inundation, and lateral subsurface flow, exert a first-order influence on the partitioning of water and energy at the land surface, yet these processes are simplified or entirely omitted in coarse-resolution LSMs (Wood et al., 2011). Vergopolan et al. (2022) showed that 80% of the spatial information on soil moisture is lost across the Contiguous United States (CONUS) when soil moisture is progressively coarsened from 30 m to 1 km. Similarly, Li et al. (2024) found that 31%–54% of spatial information in soil moisture, latent heat, and radiation fields is lost when upscaling from 1 to 12 km. Resolving these processes at k-scale is therefore essential for improving the fidelity of simulated terrestrial water, energy, and carbon cycles, as well as their coupling to the atmosphere (Wood et al., 2011; Vergopolan et al., 2022).

A prerequisite for k-scale LSM simulations is higher resolution datasets for parameterizing, calibrating, and validating model results, and access to such k-scale datasets is becoming routine. Li et al. (2024) recently developed a global 1 km dataset of land surface parameters that includes land use and land cover, vegetation, soils, and topography parameters for supporting k-scale LSMs. Rasmussen et al. (2023) has developed CONUS404, a 4 km, 40 year WRF-based hydroclimate reanalysis over CONUS by dynamically downscaling ERA5 over the CONUS to provide self-consistent, high-resolution forcing for land surface modeling. High-resolution observational and observation-constrained datasets are also increasingly available for calibrating uncertain model parameters. Examples include a 1 km global water table depth (WTD) dataset developed by combining more than 1.6 million well observations with groundwater-flow modeling (Fan et al., 2013) and a 30 m global surface water dynamics dataset derived from Landsat imagery (Pickens et al., 2020). In parallel, advances in surrogate modeling, sensitivity analysis, and machine learning are making it increasingly practical to use such datasets to calibrate multiple uncertain LSM parameters (Xu et al., 2022; Fang et al., 2025). A growing collection of independent, high-resolution k-scale remote sensing and reanalysis products also enables evaluation of multiple components of the terrestrial system (Barrett, 2003; Running et al., 2017; Zhang et al., 2022; Zheng et al., 2023). Together, these developments provide the parameter, forcing, calibration, and evaluation datasets needed to develop and rigorously assess k-scale LSM simulations.

The recent advances in high performance computing are beginning to enable k-scale Earth System Model (ESM) simulations with the terrestrial components uncoupled or coupled to the atmospheric component. Hyperresolution modeling was recognized early on to require massively parallel systems capable of solving problems with $\mathcal{O}(10^9)$ unknowns (Wood et al., 2011). GPUs provide significant speedup relative to CPUs, thus the use of performance-portable programming infrastructures for all components of ESMs is being actively pursued. In atmosphere, the Simple Cloud-Resolving E3SM Atmosphere Model (SCREAM) performs global 3.25 km convection-permitting simulations using a C++/Kokkos design for exascale portability (Donahue et al., 2024). For the ocean, a GPU-based dynamical core has been successfully used to perform decade long simulation of near-global, eddy-resolving ocean dynamics per wall-clock day at 8 km resolution on only 64 GPUs (Silvestri et al., 2025). More recently, for groundwater and surface water, Kollet et al. (2025) performed a global 1 km, three-dimensional variably saturated ParFlow simulation on 320 GPUs, while for river and flood modeling, the GPU-capable two-dimensional shallow-water model RDycore achieved up to a 42× speedup over CPUs for a 471-million-cell problem within E3SM (Bisht

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

et al., 2025). Collectively, these advances indicate that the computational barriers to routine k-scale land surface and Earth system simulations are rapidly being overcome (Wang et al., 2026). However, a large-scale, multi-year k-scale LSM simulation has not yet been demonstrated, and more fundamentally, the predictive skill of LSMs at k-scale remains largely unknown.

This work aims to evaluate and improve the performance of the E3SM Land Model version 2 (ELMv2) at k-scale over the CONUS. Offline ELMv2 simulations with improved treatments for the sub-grid topographical effects on solar radiation (Hao et al., 2021) and pluvial inundation (Xu et al., 2024b) were performed for 2001–2019. The ELM surface dataset for the CONUS domain was generated using the newly developed 1 km global surface parameters (Li et al., 2024), yielding approximately 12 million active land grid cells. The simulations were driven by NLDAS2 atmospheric forcing at $0.125° \times 0.125°$ resolution. Two model parameters were calibrated to improve ELMv2's simulation of WTD and inundated surface water fraction (f_{ssw}): a subsurface drainage parameter, f_{drain}, following the approach of Bisht et al. (2018), and a pluvial flood dynamics parameter, f_c, following Xu et al. (2024b). An additional ten datasets were used to independently evaluate the performance of simulate water, snow, near-surface energy, and carbon dynamics using calibrated ELM (ELM_{CAL}). To isolate the effect of parameter calibration, we also evaluated ELM simulation with default ELM parameters (ELM_{DEF}) against the benchmarks. Additionally, using the approach of Vergopolan et al. (2022) and Li et al. (2024), we computed the information loss for energy fluxes and soil moisture when upscaling the simulation results from a 1 km simulation to 12 km.

2 Model Description

The ELM version 2 (ELMv2) is the land component of E3SM version 2 and simulates terrestrial biogeophysical and biogeochemical processes (Golaz et al., 2022). Typical spatial resolution of ELM in century-long E3SM simulations ranges from 165 km in the low-resolution configuration (Golaz et al., 2019, 2022) to 25 km in the regionally refined North America configuration (Tang et al., 2023). ELMv2, which originated from the Community Land Model 4.5 (Oleson et al., 2010), can be configured in the following two ways: (i) Satellite Phenology (SP) mode, in which the leaf and stem area indices are prescribed, or (ii) Biogeochemistry (BGC) mode, in which vegetation phenology and soil BGC are prognostic. The two new parameterizations developed after the release of ELMv2 and included in this study address the impacts of sub-grid topography on solar radiation and pluvial inundation. These parameterizations are briefly described here.

ELMv2 assumes the Earth's surface is flat and uses a plane-parallel radiative transfer scheme based on a two-stream approximation (Dickinson, 1983). A parameterization for direct and diffuse solar radiation that accounts for sub-grid topography, based on multiple linear regression, was recently implemented in ELM (Hao et al., 2021). The four additional parameters required for the sub-grid parameterization include slope, aspect, sky view factor, and terrain configuration factor; all four variables can be easily derived from sub-kilometer scale global DEMs. The sub-grid parameterization does not modify the incoming solar radiation; instead, it updates the direct and diffuse surface albedo over mountainous terrain. Hao et al. (2021) reported that the inclusion of the sub-grid parameterization led to differences in net radiation that were as large as 20 W m^{-2} in ELM simulations with a 12.5 km spatial resolution over the Tibetan Plateau, and such differences are expected to be even greater in k-scale simulations (Hao et al., 2022; Ma et al., 2024).

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

ELMv2 allows a grid cell to be fractionally covered with standing surface water and a fraction of the first soil layer to be saturated. In the default configuration, ELM assumes that the fraction of the first soil layer that is saturated is homogeneously distributed beneath both the inundated and non-inundated portions of the grid cell. Xu et al. (2024b) showed that this default configuration of ELMv2 leads to underestimation of f_{ssw} when compared against satellite observations over North America. By modifying the default ELMv2 configuration to assume that saturation in the first soil layer is preferentially located beneath the standing surface water fraction, and by calibrating two parameters related to pluvial inundation dynamics, Xu et al. (2024b) significantly improved ELMv2's simulation of f_{ssw}.

3 Simulation Configuration

The 1 km ELMv2 surface dataset developed by Li et al. (2024), which was generated using the 1 km or finer global surface parameters, was used to perform k-scale ELM simulations. The total number of active land grid cells was 12,036,781, which is substantially greater than the number of active land grid cells in two E3SM global configurations: a 165 km resolution configuration (7,688 grid cells) and a regional refinement configuration over North America (25,055 grid cells). The BGC mode in ELMv2 requires a model spin-up simulation of 500–1000 years for climate conditions corresponding to 1850, followed by a transient simulation from 1850 to the present day. Instead, we used the SP mode configuration to make the computational cost of the k-scale simulations manageable. Hourly meteorological forcing from the North American Land Data Assimilation System Phase 2 (NLDAS2) at $0.125^\circ \times 0.125^\circ$ resolution was used to drive ELMv2 simulations from 2001 to 2019. The coarser resolution NLDAS2 forcing is spatially interpolated to the 1 km ELM grid cells via the bilinear interpolation method. The simulations were performed for 38 years by recycling 19 years of NLDAS2 atmospheric forcing. The first 19 years were considered model spin-up and were excluded from model calibration and analysis.

To improve ELM-simulated WTD, we calibrated a subsurface drainage parameter, f_{drain}, following the approach developed by Bisht et al. (2018). This approach first required six simulations with spatially homogeneous values of f_{drain} of 0.2, 0.3, 1.0, 2.5, 5.0, and 10.0. Next, a nonlinear relationship was fitted for each grid cell between the 19-year average WTD and f_{drain}. Finally, an optimal f_{drain} was computed from the fitted relationship to best match the benchmark WTD dataset described in Section 4.1.

After f_{drain} calibration, we calibrated a pluvial flood dynamics parameter, f_c, to improve ELMv2's simulation of f_{ssw} by following the approach of Xu et al. (2024b). Thus, an additional set of five ELMv2 simulations were performed with spatially homogeneous f_c values of 0.01, 0.1, 0.2, 0.3, and 0.4. A linear relationship was fitted between f_c and f_{ssw} for each grid cell, and an optimal f_c was estimated to match the benchmark f_{ssw} dataset, which is described in Section 4.2. Finally, a 38-year simulation with calibrated f_{drain} and f_c was performed, and outputs at a monthly timescale were archived for model evaluation.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

4 Datasets and Evaluation

Twelve datasets were used in this study for model calibration and evaluation. The WTD and surface water dynamics datasets were used for model calibration, while the remaining ten independent datasets were used to evaluate ELM-simulated water, snow, near-surface energy, and carbon-cycle dynamics. Datasets finer than 1 km were upscaled to ELMv2's 1 km grid, while for datasets coarser than 1 km, ELM results were interpolated to the datasets' resolution, and both datasets and model results were temporally aggregated to a monthly timescale. A brief description of all datasets used in the study is provided below.

4.1 Water Table Depth

Fan et al. (2013) developed a WTD dataset at 1 km resolution by combining more than 1.6 million well observations from North America, Western Europe, Australia, South America, Africa, and Asia with a groundwater flow model that accounts for lateral subsurface flow. The benchmark dataset shows that the WTD distribution peaks at shallow water tables and decreases exponentially for deeper WTDs. The WTD pattern at global, regional, and local scales is controlled primarily by proximity to the ocean, climatic conditions, and topography, respectively. The dataset provides a climatological average of WTD and does not include transient dynamics. It also does not include the impacts of groundwater pumping, irrigation, and drainage. Despite these limitations, this dataset has been widely used to validate equilibrium WTD simulated by LSMs (Bisht et al., 2018; Xie et al., 2020; Lv et al., 2021).

4.2 Surface Water Fraction

The global surface water dataset from the Global Land Analysis & Discovery (GLAD) (Pickens et al., 2020) was used as a benchmark for evaluating ELM-simulated surface water dynamics. The GLAD datasets are derived at 30 m spatial resolution from Landsat 5, 7, and 8 images from 1999–2020 and are available at monthly, yearly, and seasonal timescales. The dataset used a sample-based area estimator for water extent that additionally provides an estimate of uncertainty, which cannot be obtained from a pixel-counting method. The GLAD dataset was upscaled to 1 km following Xu et al. (2024b), and only the annual dataset was used for model evaluation because there are many gaps in the monthly data. Rivers, lakes, and reservoirs were removed from the GLAD dataset to be consistent with the representation of surface water in ELM. Unlike previous studies that used GLAD to evaluate the total inundation simulated by the land and river components of E3SM (Xu et al., 2024a, b), here we compare only land-simulated inundation against GLAD.

4.3 Soil Moisture

We used the recently developed 1 km gap-free global daily surface soil moisture dataset of Zheng et al. (2023) to evaluate the thickness-weighted average soil moisture in the first two soil layers of ELM. This dataset first applies a gap-filling algorithm to a 0.25° surface soil moisture product and then downscales the gap-filled data to 1 km using a random forest algorithm. The dataset has been validated to show good agreement with in situ observations from the International Soil Moisture Network and other remote-sensing products.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

4.4 Runoff

The Global Reach-level Flood Reanalysis (GRFR) is a model-based, 3-hourly, global runoff and river discharge dataset at 0.05° spatial resolution spanning 1980–2019 (Yang et al., 2021). GRFR runoff is modeled using the Variable Infiltration Capacity model, which uses precipitation forcing from the Multi-Source Weighted-Ensemble Precipitation dataset (Beck et al., 2019) and all other meteorological fields from the European Center for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis (ERA5) (Hersbach et al., 2018). GRFR runoff was bias-corrected based on percentile matching with in-situ streamflow measurements. In this study, we used the monthly GRFR runoff dataset to evaluate ELM-simulated runoff, and ELM runoff was upscaled to 5 km to match the resolution of GRFR.

4.5 Snow Dynamics

Following Hao et al. (2023), we used in situ observations and data-assimilated products to evaluate snow dynamics simulated by ELM. Specifically, we used the daily Bias Correction and Quality Control (BCQC) snow water equivalent (SWE) dataset from 788 Snow Telemetry (SNOTEL) sites. We also used two data-assimilated SWE products: the 4 km daily University of Arizona (UA) snow product, version 1 (Zeng et al., 2018), and the 1 km daily SNOw Data Assimilation System (SNODAS) product developed by the NOAA National Weather Service's National Operational Hydrologic Remote Sensing Center (Barrett, 2003).

4.6 Latent Heat Flux

The Moderate Resolution Imaging Spectroradiometer (MODIS) derived latent heat flux (LH) product, MOD16A2GF 6.1, was used for evaluating ELM-simulated LH. MOD16A2GF is a gap-filled 8-day composite product with a 500 m spatial resolution that is generated at the end of a year (Running et al., 2017). The dataset estimates evapotranspiration using the Penman–Monteith equation, along with other MODIS-derived vegetation datasets and daily surface meteorological inputs from the Modern-Era Retrospective analysis for Research and Applications (Mu et al., 2011).

4.7 Land Surface Temperature

The 1 km daily Land Surface Temperature (LST) dataset developed by Zhang et al. (2022) is a seamless, spatiotemporally gap-filled global product. It includes midday (13:30) and midnight (01:30) LST values that are consistent with the MODIS Terra (MOD11A1) and Aqua (MYD11A1) products. The accuracy of this gap-filled LST product is comparable to or better than that of other seamless LST datasets (Zhang et al., 2022). We used the average of the midday and midnight LST values to evaluate daily ELM-simulated LST.

4.8 Gross Primary Productivity

Three GPP datasets were used to evaluate ELM performance. The first dataset was the 8-day MODIS MOD17A2H product at 500 m resolution (Running et al., 2017). This dataset has been used extensively to evaluate land surface models (Collier et al., 2018; Lawrence et al., 2019; Li et al., 2026). The second dataset was the monthly FLUXCOM-X product at 0.05° resolution,

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

which upscales eddy-covariance observations using machine learning with satellite and meteorological inputs and represents a more advanced and flexible next-generation framework than earlier FLUXCOM products (Nelson et al., 2024). The third dataset was an 8-day GPP product at 500 m resolution based on a biophysical model that uses MODIS datasets and Global Land Data Assimilation System (GLDAS) atmospheric forcing (Zhang et al., 2019). This dataset has been shown to perform similarly to or better than other state-of-the-art products in a comparison against 95 global tower sites (Zhang et al., 2019).

4.9 Evaluation and Analysis

To quantitatively assess ELMv2 performance against the benchmark datasets, we employed five metrics: bias, relative bias, root-mean-square error (RMSE), relative RMSE (rRMSE), and the coefficient of determination (R^2). For a set of n paired model (m_i) and benchmark (b_i) values, these metrics are defined as follows. The bias quantifies the mean systematic difference between the model and the benchmark,

$$\text{Bias} = \frac{1}{n}\sum_{i=1}^{n}(m_i - b_i), \tag{1}$$

and the relative bias normalizes this quantity by the mean of the benchmark,

$$\text{Relative Bias} = \frac{\sum_{i=1}^{n}(m_i - b_i)}{\sum_{i=1}^{n} b_i} \times 100\%. \tag{2}$$

The RMSE measures the overall magnitude of the error,

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - b_i)^2}, \tag{3}$$

with rRMSE computed as

$$\text{rRMSE} = \frac{\text{RMSE}}{\frac{1}{n}\sum_{i=1}^{n} b_i} \times 100\%, \tag{4}$$

The coefficient of determination (R^2) measures the strength of the linear association between the model and the benchmark,

$$R^2 = \left(\frac{\sum_{i=1}^{n}(m_i - \bar{m})\left(b_i - \bar{b}\right)}{\sqrt{\sum_{i=1}^{n}(m_i - \bar{m})^2}\sqrt{\sum_{i=1}^{n}\left(b_i - \bar{b}\right)^2}} \right)^2, \tag{5}$$

where $\bar{m}$ and $\bar{b}$ is the mean of the model and benchmark values, respectively. These metrics were computed both spatially, across grid cells for temporally-averaged fields, and temporally, across the monthly time series of spatially-averaged values, following the approaches of Hao et al. (2023) and Li et al. (2026).

In addition, we quantified the information loss (γ) incurred when representing high-resolution fields at coarser spatial scales, following the procedure described in Li et al. (2024). Specifically, the 1 km land ELM-simulated variables were upscaled to 12 km and the spatial standard deviation (σ) was calculated within each $0.5^0 \times 0.5^0$ box from the original 1 km (σ^{1k}) values and from the upscaled 12 km (σ^{12k}) values. Then, information loss was computed as

$$\gamma = \left(1 - \frac{\sigma^{12k}}{\sigma^{1k}}\right) \times 100\% \tag{6}$$

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

5 Results

Two simulations were performed that differ only in the values of f_{drain} and f_c: ELM_{DEF} uses the default values ($f_{drain} = 2.5$, $f_c = 0.4$) and ELM_{CAL} uses the calibrated values described in Section 3. Forcing, surface data, and spin-up are otherwise identical. The evaluation statistics for ELM_{CAL} presented in Tables 3. A similar evaluation ELM_{CAL} is instead presented in Supplementary Material because the results are not significantly different than ELM_{CAL}. Unless noted, the spatial maps shown below are from ELM_{CAL}; differences from ELM_{DEF} are quantified in the text.

5.1 Water table depth

The WTD benchmark of Fan et al. (2013) shows a large spatial variability across the CONUS domain, with deeper WTD in mountainous regions and shallower WTD in riparian and lowland areas (Figure 1a). The estimated f_{drain} values exhibit a spatial pattern closely aligned with the benchmark, with lower parameter values in regions with deeper WTD and vice versa (Figure 1b). An optimal value of f_{drain} was identified for 55% of the grid cells within the study domain using the six ensemble simulations. For the remaining grid cells, the WTD benchmark fell outside the range simulated by the ELM ensembles, which produced WTDs in all six f_{drain} values deeper than the benchmark for 26% of the grid cells and shallower for 19%. The ELM simulations using the default f_{drain} show limited spatial variability in WTD compared to the benchmark (Figure 1c), while the use of calibrated f_{drain} produces noticeably better agreement (Figure 1d). The WTD simulated with calibrated f_{drain} shows substantial improvement, with a reduced bias (from -27.6 m to -7.3 m), a reduced RMSE (from 54.7 m to 32.8 m) and an increased R^2 (from 0.27 to 0.80), as summarized in Table 2. In the current ELM structure, the maximum WTD depth is limited to 80 m. When the comparison between the benchmark and ELM simulations is restricted to grid cells with benchmark WTD shallower than 80 m, the agreement improves further across all three statistical metrics (second row of Table 2).

In order to gain a better perspective of the details included within this simulation relative to a typical global Earth System Model with a standard resolution, we zoom in over two regions with a $1^0 \times 1^0$ bounding box (Figure 2). These two regions were selected as they approximately correspond to the regions presented in the recent Ma et al. (2026) study, which was focused on developing sub-kilometer groundwater estimates using machine learning approaches. The higher estimated f_{drain} values correspond to shallower WTD for both regions. The ELM simulation using calibrated f_{drain} accurately captures the spatial variability of WTD. However, for the portions of the two regions with shallower WTD in the benchmark (as shown with dark blue color), the ELM simulation predicts a deeper WTD (as shown with a lighter blue color).

5.2 Surface water dynamics

The upscaled 19-year average surface water fraction from GLAD shows higher values along major river networks and in low-lying coastal areas (Figure 3a). Using the default ELM value of $f_c (= 0.04)$ leads to a substantial overestimation of simulated f_{ssw} across the central and eastern CONUS with a bias, RMSE, and R2 of 3.0×10^{-2}, 5.3×10^{-2}, and 0.14, respectively (Table 2). When the calibrated f_c values are used (Figure 3b), ELM more accurately captures surface water dynamics spatial heterogeneity (Figure 3d) by reducing the bias by two orders of magnitude, lowering the RMSE by 2.6 times, and increasing

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

R2 by 4.4 times (Table 2). Although calibration significantly improves the performance of f_{ssw}, ELM overestimates it on most of the study domain (Figure 3f).

5.3 Soil moisture

The Zheng2023 benchmark shows a CONUS-scale gradient with a drier Southwest and Great Plains and a wetter central and east CONUS (Figure 4a). ELM is able to accurately capture this large-scale structure well (Figure 4b) and has a relative bias, rRMSE, and R2 of 16.5%, 24.3%, and 0.83, respectively. The bias is largest over the Midwest, Northeast, and Northwest (Figure 4c). Across all months, ELM simulates a higher soil moisture as compared to the benchmark (Figure 4d). The overestimation of soil moisture by ELM is consistent with the recent findings of Li et al. (2026), although that study only focused on comparing ELM simulations for the Mid-Atlantic region. While ELM and Zheng2023 agree on the timing of lowest soil moisture during the summer, the amplitude of of the seasonal change is much larger in ELM. Model calibration slightly increases the wet bias in near-surface soil moisture against Zheng2023 (spatial relative bias $+17.3\%$ to $+19.3\%$; temporal $+10.5\%$ to $+12.4\%$), while RMSE (0.06 and 0.03) and correlations (0.83 spatial, 0.98–0.99 temporal) are unchanged.

5.4 Runoff

The annual-average total runoff in the GRFR dataset shows low values for the Great Plains, a sharp transition zone through western plains, and high values for the Midwest, Southwest, Northeast and mountainous regions of the Pacific Northwest (Figure 5a). The bias of ELM-simulated runoff relative to the GRFR dataset has a spatial pattern that follows the spatial variability in the GRFR dataset, with the model underestimating regions of high runoff and overestimating regions of low runoff (Figure 5c). The model's inability to accurately capture the spatial minima and maxima also impacts simulation of monthly runoff values (Figure 5d). The monthly simulated runoff averaged over the CONUS is significantly muted as compared to the GRFR dataset. Moreover, ELM's runoff peaks a month earlier in March relative to the GRFR dataset, while the lowest runoff is predicted three months later in November. Relative to ELM_{DEF}, calibration reduces the spatial RMSE of annual runoff against GRFR from 147.6 to 142.9 mm yr^{-1} (rRMSE 60.8% to 58.8%) and increases the temporal correlation from 0.55 to 0.57, at the cost of a slightly larger temporal bias (-9.7% to -10.5%); the seasonal cycles of the two simulations are otherwise nearly indistinguishable (Figure 5d).

5.5 Snow water equivalent and snow depth

SNODAS and UA SWE benchmarks agree on the spatial distribution with higher values in the mountainous area (Figure 6a,b), though they differ in their magnitudes. After excluding regions with no SWE values, the SNODAS benchmark has a lower mean SWE of 12.5 mm as compared to the UA benchmark, which has a mean value of 39.2 mm. While ELM is able to capture the spatial distribution of SWE, the mean SWE simulated by ELM is much lower than both benchmarks, with a mean of 6.5 mm for snow-covered regions. ELM has a mean bias of -0.6×10^{-3} m both against SNODAS and UA benchmarks. These results are consistent with the previous finding of Hao et al. (2023) that also observed a similar underprediction of SWE by ELM

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

over the western United States. Compared to the SNOTEL stations, ELM has a larger underprediction of SWE with a bias of -1.2×10^{-1} m, which is 20 times higher compared to SNODAS and UA. The model has poorer performance against SNOTEL sites in the Cascade Range with higher bias and lower R2 (Figure 7), which is also consistent with prior results (Hao et al., 2023). Since SNOTEL data had many monthly missing values, we did not perform a temporal evaluation of ELM simulation against SNOTEL data. The comparison of ELM-simulated snow depth against SNODAS and UA benchmarks shows spatio-temporal results that are similar to SWE results, with the model similarly underestimating snow depth over the mountainous area (Figure S1). The relative bias spatially and temporally is less than 1% for both benchmarks and R^2 values are higher than 0.8 (see Table 3). SWE and snow depth are insensitive to calibration with relative biases against SNODAS, UA, and SNOTEL differing by less than 1.5% between ELM_{CAL} and ELM_{DEF}, and spatial and temporal correlations remaining identical.

5.6 Latent Heat Flux

The MODIS LH dataset shows a transition from higher LH values over the eastern CONUS to lower LH values for the western CONUS (Figure 8a). While ELM is able to capture the same east-west transition (Figure 8b), it predicts significantly higher LH over Texas, Kansas, and Oklahoma (Figure 8c). ELM simulated higher maximum LH and lower minimum LH value relative to the MODIS dataset (Figure 8d). ELM predicts 37% higher LH value averaged over June-July-August. Additionally, the timing of peak LH values between ELM and MODIS does not match, with ELM peaking a month earlier in June. The spatio-temporal relative bias ranges between 19.3% and 21.3% and is lower than the rRMSE, which ranges from 31.6% to 40.5%, while the R2 between ELM and MODIS remains high at 0.89 to 0.95 (Table 3). Calibration slightly increases the LH overestimation against MODIS (spatial bias $+15.3\%$ to $+19.3\%$; temporal rRMSE 34.9% to 40.5%), while spatial and temporal correlations are unchanged (0.88–0.89 and 0.95–0.96), indicating that the summertime LH bias is not a consequence of the parameter calibration.

5.7 Land Surface Temperature

The MODIS LST dataset shows a north-south gradient with cooler surface temperature for the northern states and mountainous western regions, and warmer southern states (Figure 9a). It should be noted that the MODIS product is based on four overpass snapshots, while ELM values are averages from 30-minute time steps. Despite differences in temporal sampling, ELM is able to accurately capture the observed spatial gradient in temperature with a mean bias less than 1K (Figure 9b). The seasonal variation in LST as predicted by ELM agrees remarkably well, as shown in Figure 9c and as quantified by all three metrics summarized in Table 3. The peak summer temperature, as well as cool winter temperature, are accurately predicted by ELM. The LST statistics also essentially remained unchanged by calibration with ELM_{CAL} and ELM_{DEF} both have a similar spatial bias of about -0.75 K, a spatial RMSE of 1.88–1.89 K, and identical spatial (0.96) and temporal (1.00) correlations against MODIS benchmark.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

5.8 Gross Primary Production

All three GPP benchmarks have a similar spatial gradient pattern with higher values in the east and lower values in the west (Figure 10a-c), and this east-west gradient is consistent with the MODIS LH benchmark dataset (Figure 8a). The spatially averaged GPP values for MODIS, FLUXCOM-X and Zhang2019 were 2.15 [$\mu CO_2 m^{-2} s^{-1}$], 2.43 [$\mu CO_2 m^{-2} s^{-1}$], and 2.83 [$\mu CO_2 m^{-2} s^{-1}$], respectively; while the ELM simulated GPP was 2.62 [$\mu mol CO_2 m^{-2} s^{-1}$]. ELM accurately captures the east-west gradient in GPP (not shown), though it overpredicted GPP relative to benchmarks for central and eastern CONUS and underpredicted GPP for the southwestern domain. The largest bias in ELM simulation of GPP is against the MODIS benchmark, with strongest overestimation over the south-central and south-eastern states (Figure 10d). The intensity of bias against the FLUXCOM-X benchmark is smaller as compared to MODIS, while the spatial pattern remains the same (Figure 10e). ELM simulated GPP shows an overall underestimation against the Zhang2019 benchmark (Figure 10f).

The three benchmarks also have differences in the seasonal cycle of GPP (Figure 11). The MODIS benchmark peaks in June, while both FLUXCOM-X and Zhang2019 peak one month later in July. Similar to their spatially average values, the MODIS benchmark has the lowest peak GPP value, while Zhang2019 has the largest peak value. The timing of ELM's peak GPP matches that of MODIS, though the magnitude of the peak GPP is more comparable to Zhang2019. During the winter months (DJF), ELM agrees perfectly with the low values of MODIS and FLUXCOM-X; however, ELM shows a much more rapid increase during spring compared to all three benchmarks.

GPP statistics are nearly insensitive to calibration: biases against the three benchmarks change by at most 0.06 μmol CO_2 m^{-2} s^{-1} and correlations are identical (0.86–0.92 spatial, 0.93–0.97 temporal).

5.9 Information loss of ELM-simulated energy and water variables

Significant information loss with a distinct east-west gradient was observed when 1 km ELM simulations were upscaled to 12 km (Figure 12). The spatial mean value of γ for net radiation, latent heat flux, sensible heat flux, and near surface soil moisture was 43%, 41%, 44%, and 34%, respectively. The information loss for the Southeast and the Great Plains region was much smaller for LH and SM, while the same region had higher information loss for Rnet and SH.

The information loss exhibited monthly and interannual variations for all four variables (Figure 13). The information loss for SH and LH had high values during winter and reached their lowest values during summer. The Rnet exhibited an opposite pattern by achieving highest values during summer and lowest values during winter. Soil moisture had a consistently lower information loss compared to the energy fluxes throughout the year, although it reached peak values twice during the year (in February and July). All four variables exhibited coherent interannual variation and reached their lowest value in 2010. The latent heat flux had the sharpest decline/recovery before/after 2010, followed by sensible heat flux and net radiation. To better understand the interannual variation in information loss, we computed the time series of domain averaged mean, μ, and domain standard deviation, σ for all four variables using ELM simulations at 1 km. The results showed that the interannual variability of σ had a significantly higher correlation with information loss than μ (

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

6 Discussion

In this work, we sequentially calibrated two uncertain ELM parameters to improve prediction of WTD and surface water fraction, and performed a cumulative of 190 years of simulation with a study domain containing approximately 12 million grid cells over the Contiguous United States. The approach of sequential calibration of model parameters is not expected to scale for a global 1 km mesh with more than 150 million grid cells, as well as for a larger number of uncertain parameters that control the quality of the ELM hydrological simulation (Xu et al., 2022). Alternatively, a surrogate-assisted calibration offers a potential scalable solution. Xu et al. (2022) developed an inexpensive surrogate of ELM using polynomial chaos expansion machinery with Bayesian compressed sensing for 11 runoff relevant parameters, identified the three most important parameters using Sobol' index sensitivity analysis, and estimated optimal values for the most sensitive parameters to significantly improve model performance. More recently, Fang et al. (2025) used a machine learning based surrogate framework to calibrate five uncertain ELM parameters, although their results showed that improvements in the prediction of WTD did not lead to better agreement in the simulation of runoff, soil moisture, and evapotranspiration. Keetz et al. (2025) combined data assimilation and machine learning to develop the "calibrate, emulate, sample" (CES) framework for inferring six uncertain, plant-functional-type-specific photosynthetic parameters in the demographic vegetation model CLM-FATES. The CES framework combines ensemble Kalman smoothing, a machine-learning-based emulator, and adaptive Markov chain Monte Carlo sampling, thereby reducing the number of computationally expensive model simulations to a few hundred, compared with the hundreds of thousands required by traditional calibration approaches.

The limited impoact of calibration beyond WTD and surface water fraction is consistent with **?**, who found that machine-learning calibration of groundwater-related parameters in ELM improved WTD while leaving, or in some regions degrading, the performance of runoff, soil moisture, and ET, and who noted that the reasonable skill of the uncalibrated model raises questions about calibration priorities. Together, these results indicate that the remaining biases in LH, GPP, soil moisture, and snow reflect structural or forcing deficiencies that cannot be remedied by tuning drainage parameters alone, motivating the process-level improvements discussed below.

While this work used the new 1 km dataset for the land use/land cover, vegetation, soils, and topography from Li et al. (2024), the urban land unit dataset was based on the MODIS urban and built-up classification of Jackson et al. (2010) (J2010) corresponding to three density classes, i.e., tall-building district, high-density, and medium density. Recently Cheng et al. (2025) has developed a 1 km spatially continuous urban surface property dataset, called U-Surf, that derives internally consistent radiative, morphological and thermal parameters directly for the different facets of urban canyon parameterizations for LSMs, including ELM. Future 1 km ELM simulations would therefore benefit from replacing the J2010-derived urban parameters inherited in our current setup with U-Surf, which better resolves fine-scale intra- and inter-city heterogeneity in urban surface properties.

Farmani et al. (2025a) provides a pathway forward for reducing the consistent overestimation of ELM-simulated near surface soil moisture (Figure 4d) by including missing preferential flow representation. The use of a dual-permeability scheme within the Noah-MP model accelerated drainage from the macropore flow from the surface layer, shortening surface soil moisture memory (SMM), while the same rapid bypass flow reduces root-zone SMM by routing water past the root zone into deeper

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

storage. Similarly, Farmani et al. (2025b) serves as a useful example for improving ELM's muted runoff dynamics (Figure 5d). Specifically, they found that the use of the Van Genuchten retention curve outperformed Brooks-Corey/Clapp-Hornberger, which is currently used in ELM, in matching observed baseflow index. Machine-learning-derived hydraulic parameters also lead to 15-20% improvements in streamflow prediction over default lookup table-based hydraulic parameters.

Another limitation of the current work is that it only includes vertical transport of water, while previous studies have shown that neglecting lateral transport of water can have large impacts on surface energy partitioning in regions with shallower water table (Maxwell and Condon, 2016). Approaches for incorporating lateral hydrological transport within land surface models generally fall into two categories: (1) methods that route water within a single grid cell by discretizing it into multiple hydrological response units or sub-grid landscape elements (Hazenberg et al., 2015; Swenson et al., 2019; Chaney et al., 2021), and (2) methods that explicitly transport water between all neighboring grid cells (Maxwell, 2013; Bisht et al., 2017; Miura and Yoshimura, 2020). Recently Qiu et al. (2024) extended the ELM version 2.0 to include lateral flows within the unsaturated and saturated zones; however this development only supports serial computing and must be extended for parallel computing to be included for future kilometer scale ELM simulations.

An updated version of the sub-grid topographic solar parameterization of Hao et al. (2021), named fineTOP, is now available (Hao et al., 2025). Unlike previous parameterizations that only made modifications to surface albedo, fineTOP explicitly resolves the impact of fine-scale terrain features (such as slope, aspect, and shading/reflection between neighboring grid cells) on downward shortwave and longwave radiation, and surface radiative properties. These fine-scale effects become increasingly important at 1 km resolution, but are neglected in this work. Since absorbed radiation is a primary driver of surface energy partitioning and snowmelt timing, fineTOP parameterization is expected to be included in future 1 km ELM simulations. In this work, we did not include fineTOP parameterization to establish a baseline evaluation of ELM at the k-scale using only the physics options included in ELMv2. Future work will include fineTOP as well as other recent developments for comparison with ELMv2 to document their impacts on k-scale simulations.

7 Conclusions

This study evaluated the performance of ELMv2 configured at 1 km resolution across CONUS, using a newly developed 1 km surface dataset and improved parameterizations for sub-grid topographic radiation and pluvial inundation. Two uncertain model parameters, f_{drain} and f_c, were sequentially calibrated against benchmark WTD and surface water fraction datasets, substantially reducing bias and RMSE and improving R^2 for both variables relative to the simulation with default parameters. An additional ten independent datasets were used to evaluate ELM-simulated water, snow, near-surface energy, and carbon-cycle dynamics. ELM captured the broad spatial and seasonal patterns of soil moisture, runoff, SWE, latent heat flux, land surface temperature, and GPP reasonably well, though systematic biases remained, including a wet bias in soil moisture over the Midwest, Northeast and Northwest, underestimated and muted-amplitude runoff with a shifted seasonal peak, underprediction of SWE across the mountainous West, an overestimation of summertime latent heat flux over the south-central CONUS, and an overprediction of GPP relative to MODIS and FLUXCOM-X over the central and eastern domain. Upscaling the 1 km

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

simulation to 12 km resulted in substantial information loss (34–44% on average) in net radiation, latent and sensible heat flux, and near-surface soil moisture, with information loss more strongly correlated with the spatial variability than the mean of these variables. Together, these results demonstrate both the value of k-scale ELM simulations in resolving fine-scale spatial heterogeneity in water, energy, and carbon fluxes, and highlight the continued need for improved model physics, surrogate-based calibration strategies, and higher-fidelity input datasets, as discussed above, to further reduce systematic biases as ELM moves toward global k-scale application.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

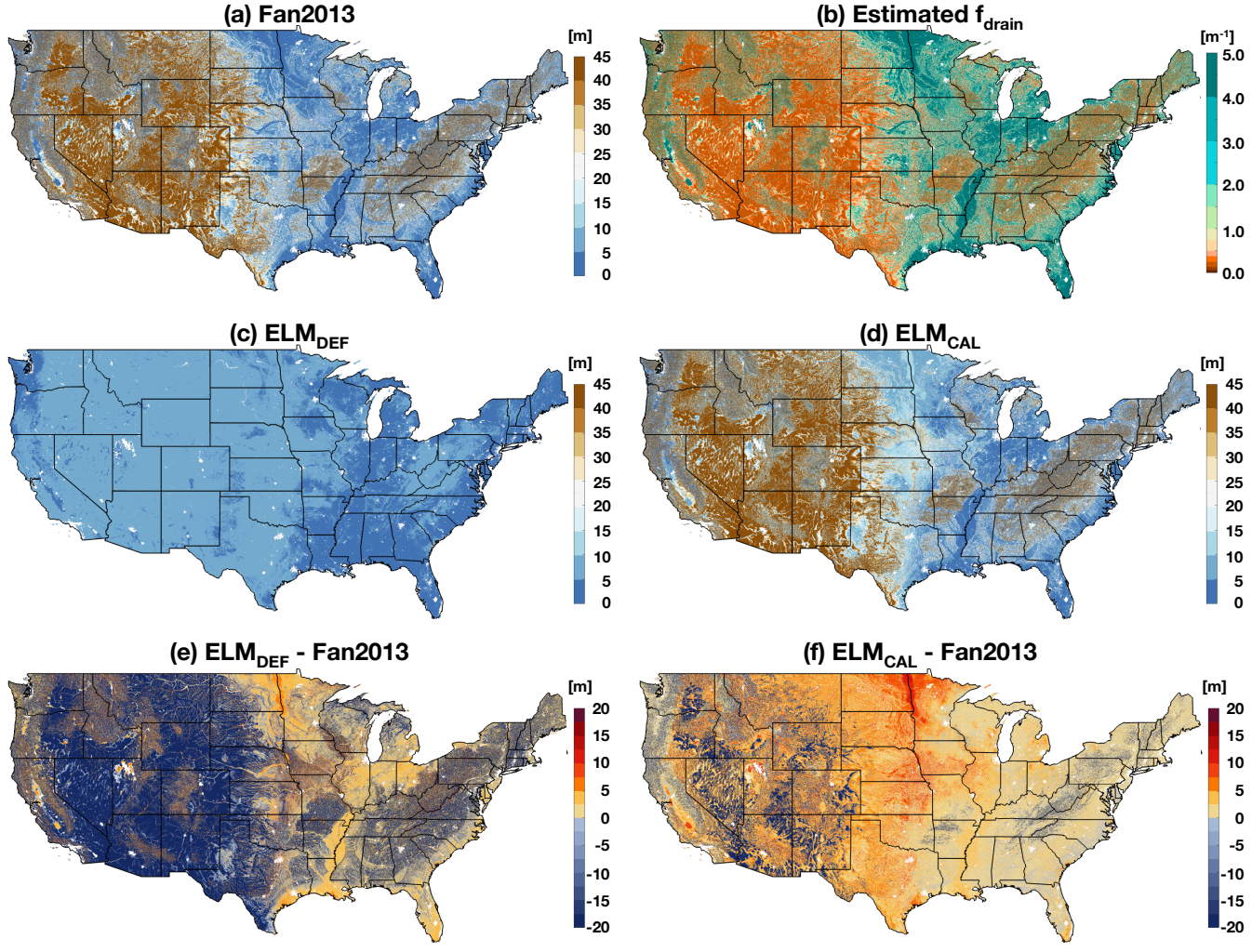


Figure 1. (a) Water table depth (WTD) benchmark of Fan et al. (2013), (b) Estimated f_{drain} parameter, (c) ELM-simulated WTD using default f_{drain}, (d) ELM-simulated WTD using estimated f_{drain}, (e) WTD bias using default f_{drain}, and (f) WTD bias using estimated f_{drain}.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

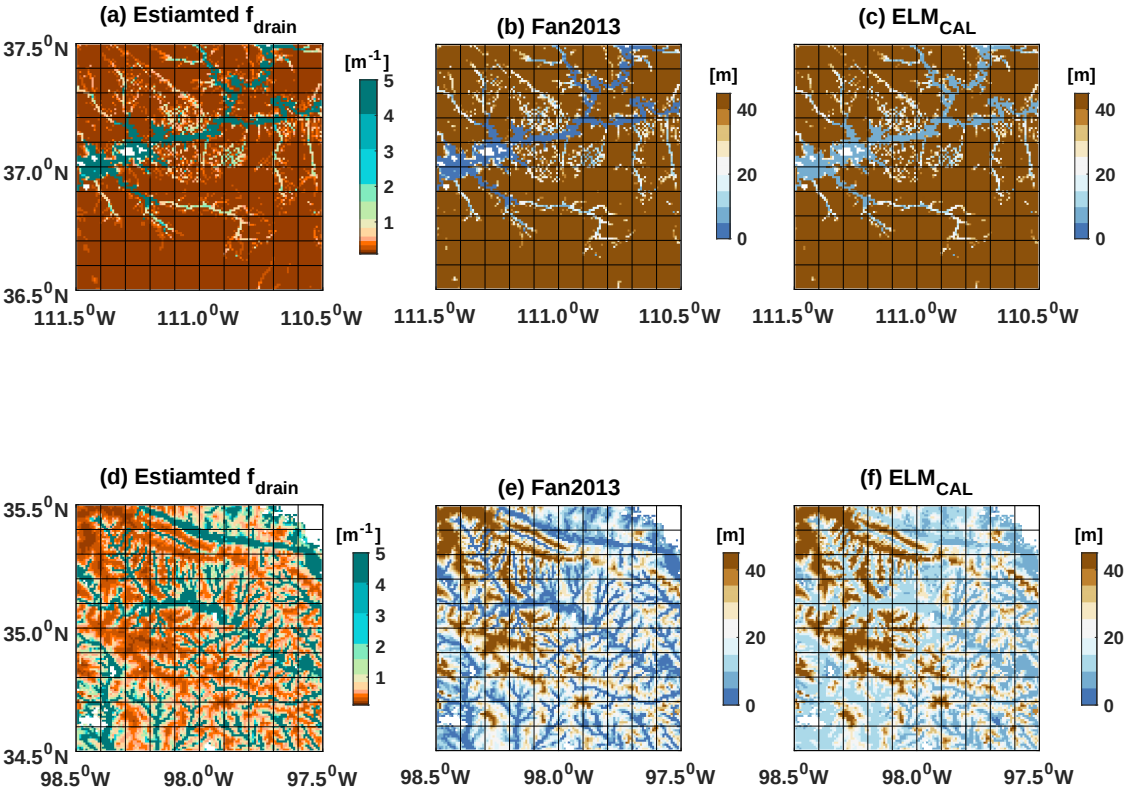


Figure 2. A zoom-in of results presented in Figure 1 for two $1^0 \times 1^0$ regions centered at (37.0^0N, 111.0^0W) in the top row and (35.0^0N, 98.0^0W) in the bottom row. (a,d) Estimated f_{drain} value for the two regions, (b,e) Water table depth (WTD) benchmark of Fan et al. (2013), and (c,f) ELM-simulated WTD using estimated f_{drain}.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

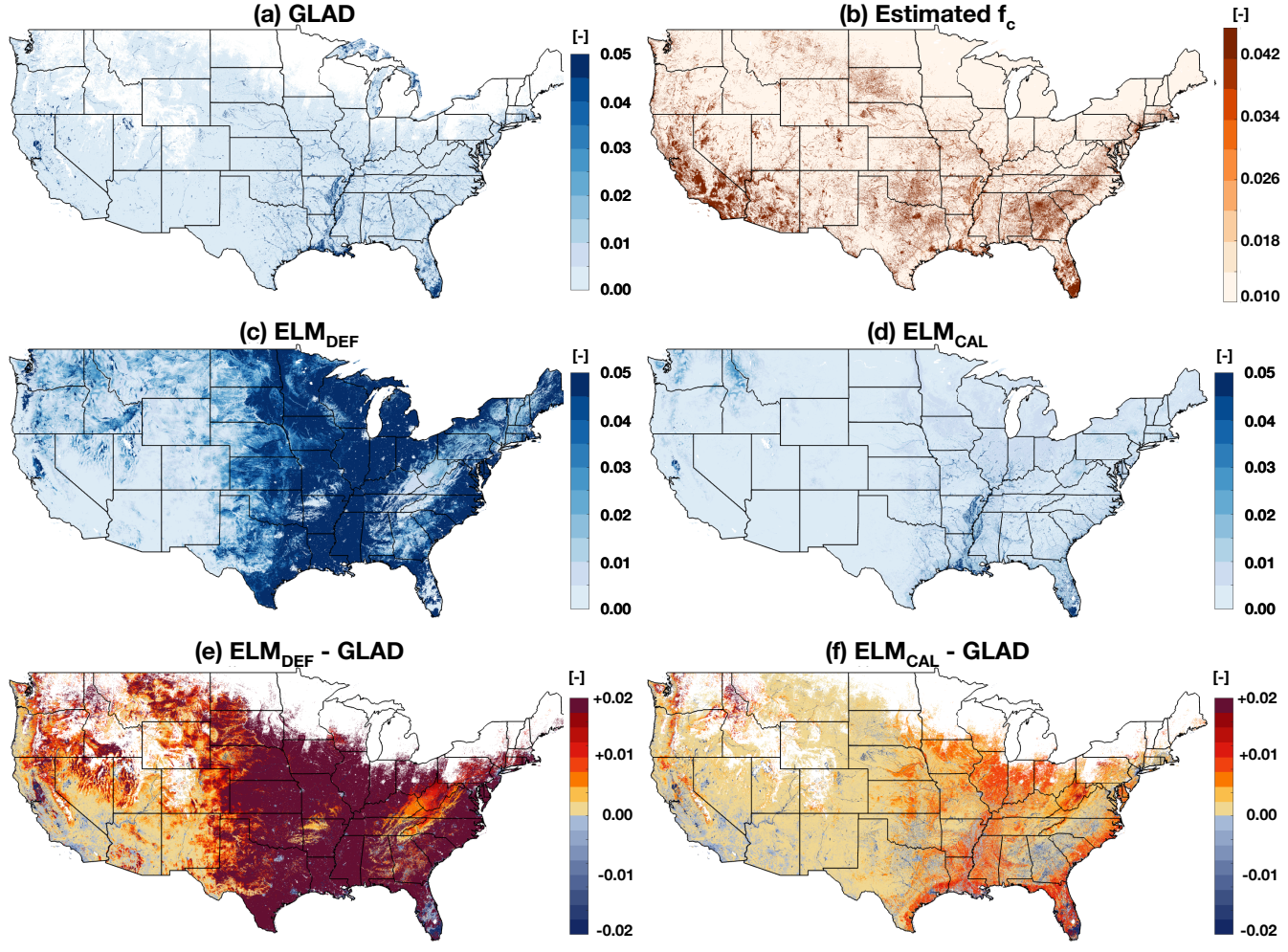


Figure 3. (a) Inland 30 m surface water dataset of GLAD of Pickens et al. (2020) upscaled to 1 km using the approach of Xu et al. (2024b), (b) Estimated f_c (c) ELM-simulated fraction of grid cell occupied by standing surface water, f_{ssw}, using default parameter value for f_c, (d) ELM-simulated f_{ssw} using a calibrated f_c value, (e) bias in the simulation of f_{ssw} using default f_c value, and (f) bias in the simulation of f_{ssw} using calibrated f_c value.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

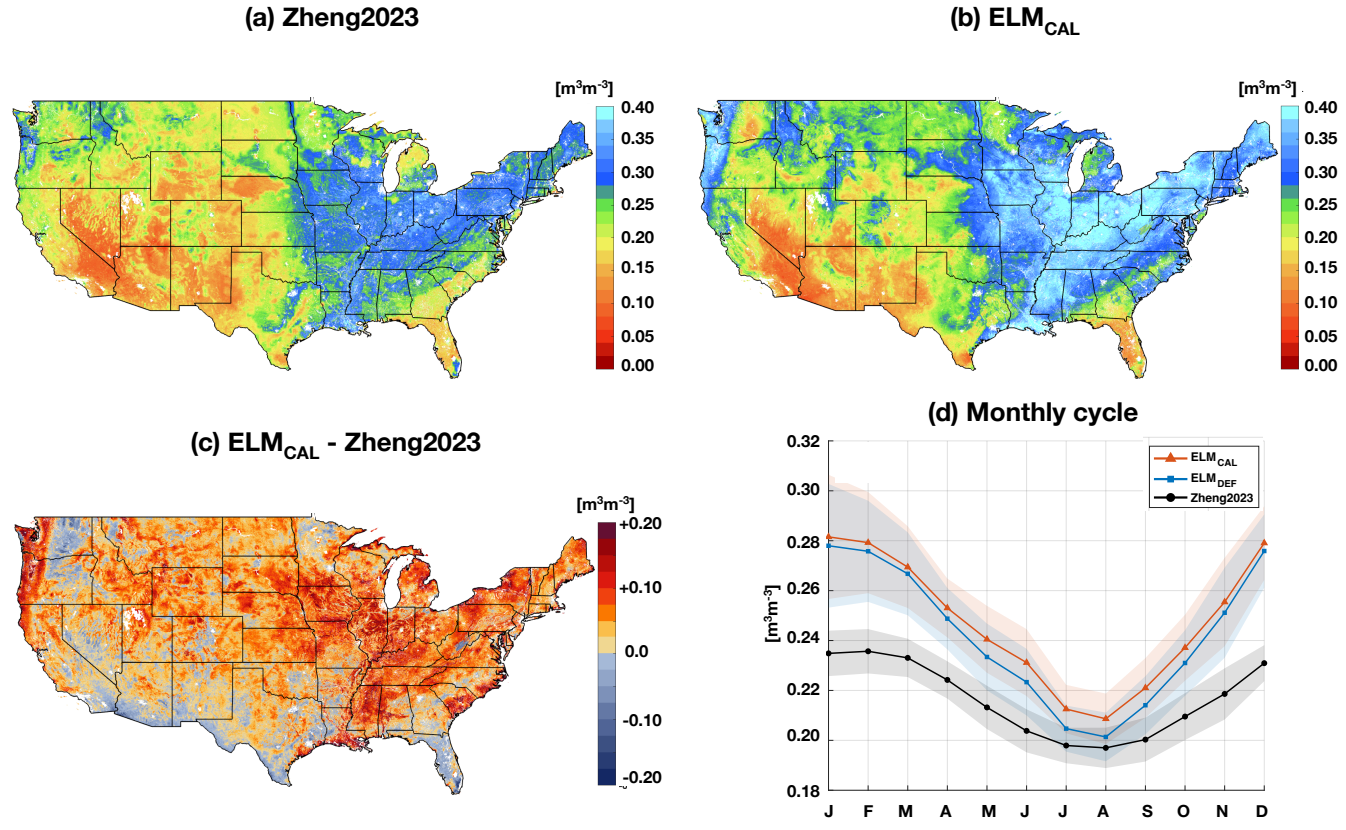


Figure 4. (a) Temporally-averaged soil moisture benchmark of Zheng et al. (2023) (Zheng2023), (b) temporally-averaged soil moisture simulated by ELM using calibrated parameters (ELM_{CAL}), (c) temporally-averaged soil moisture bias between ELM_{CAL} and Zheng2023, and (d) spatially-averaged, monthly time series of soil moisture for ELM_{CAL} and ELM using default parameters, (ELM_{DEF}) against the benchmark. The shading in the time series indicates $\pm$ one spatial standard deviation.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

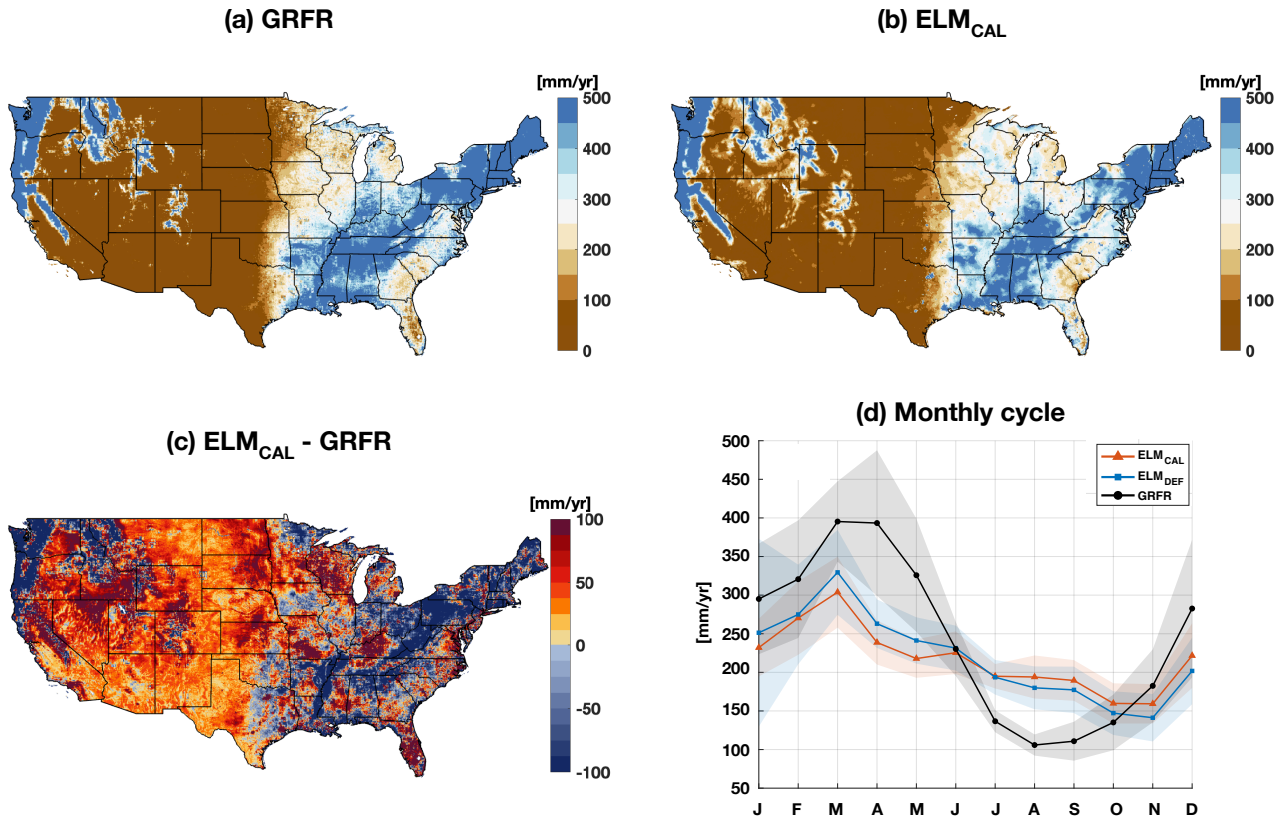


Figure 5. (a) Temporally-averaged runoff benchmark GRFR (R_{GRFR}), (b) temporally-averaged runoff simulated by ELM using calibrated parameters (ELM_{CAL}), (c) temporally-averaged runoff bias between ELM_{CAL} and GRFR , and (d) spatially-averaged, monthly time series of runoff for ELM_{CAL} and ELM using default parameters, (ELM_{DEF}) against the benchmark. The shading in the time series indicates $\pm$ one spatial standard deviation.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

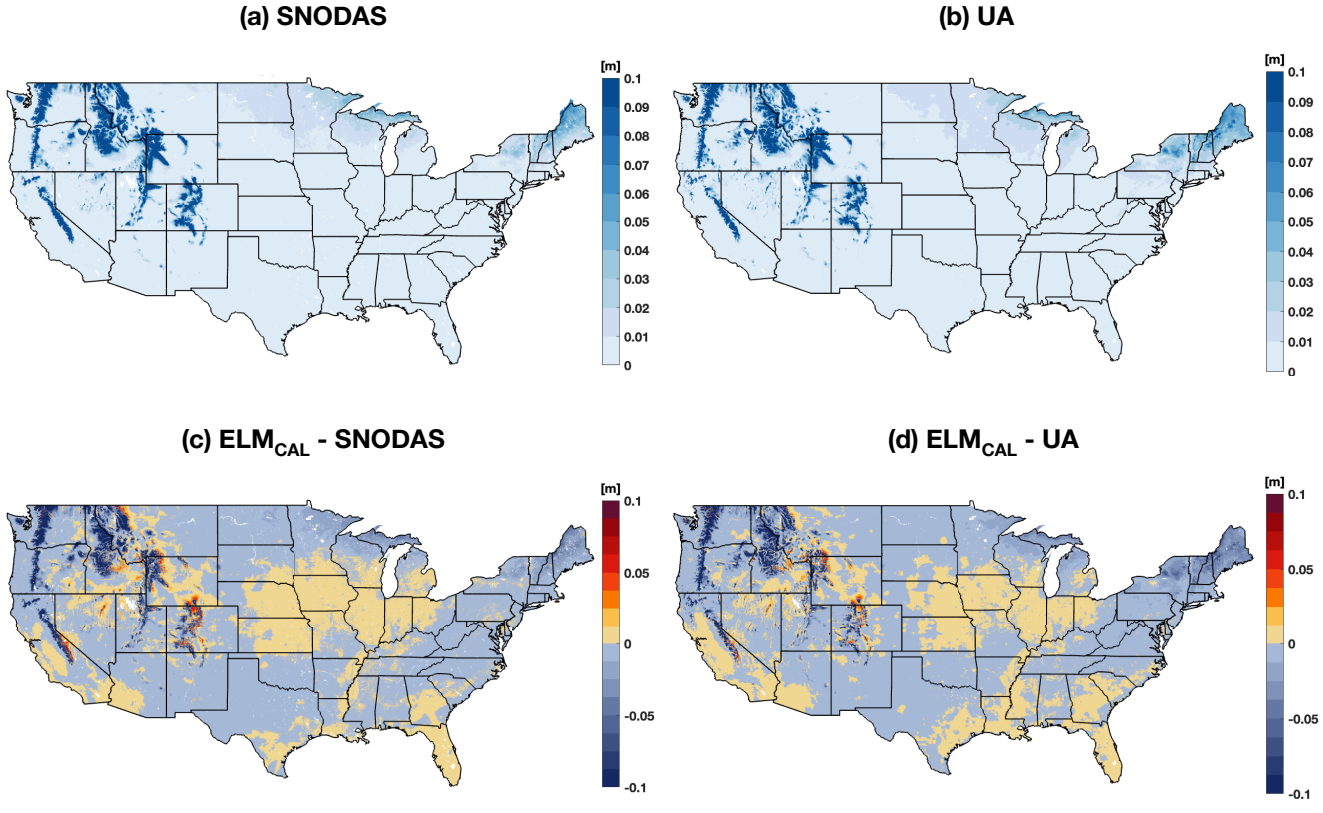


Figure 6. Temporally-averaged snow water equivalent (SWE) benchmark of (a) SNODAS at 1 km and (b) UA at 4 km. (c) Temporally-averaged bias between ELM-simulated SWE and SNODAS. (d) Temporally-averaged bias between ELM-simulated SWE and UA.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

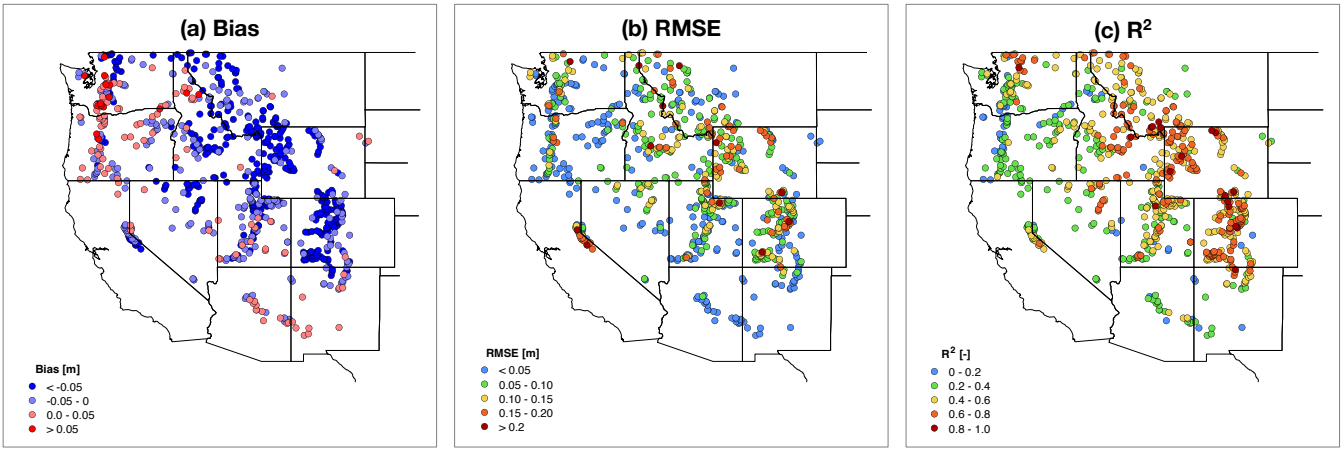


Figure 7. Spatial distribution of ELM-simulated Snow Water Equivalent (SWE) against SNOTEL benchmark for (a) bias, (b) RMSE, and (c) R2.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

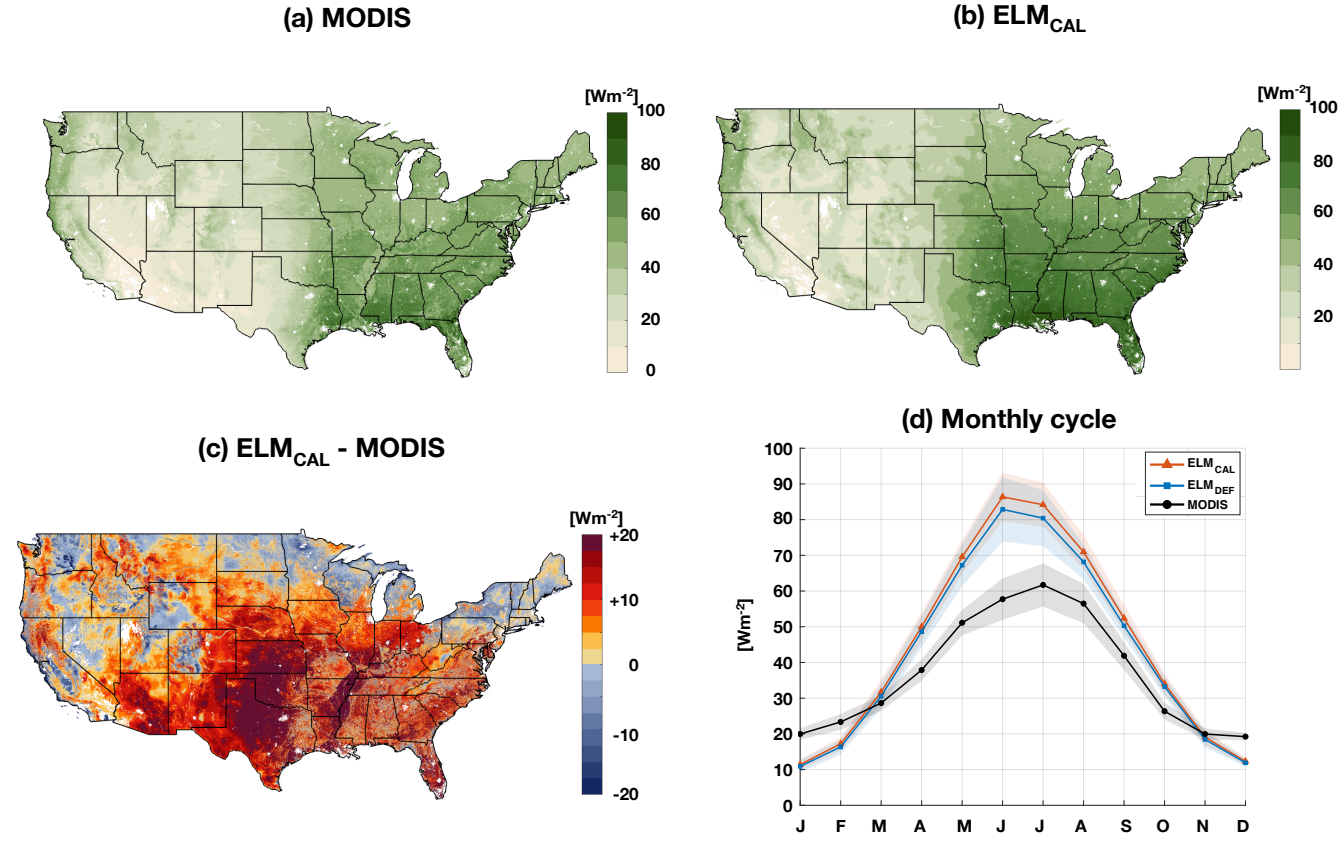


Figure 8. (a) Temporally-averaged latent heat (LH) flux benchmark MODIS (LH_{MODIS}), (b) temporally-averaged LH simulated by ELM using calibrated parameters (ELM_{CAL}), (c) temporally-averaged LH bias between ELM_{CAL} and the benchmark, and (d) spatially-averaged, monthly time series of LH for ELM_{CAL} and ELM using default parameters, (ELM_{DEF}) against the benchmark. The shading in the time series indicates $\pm$ one spatial standard deviation.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

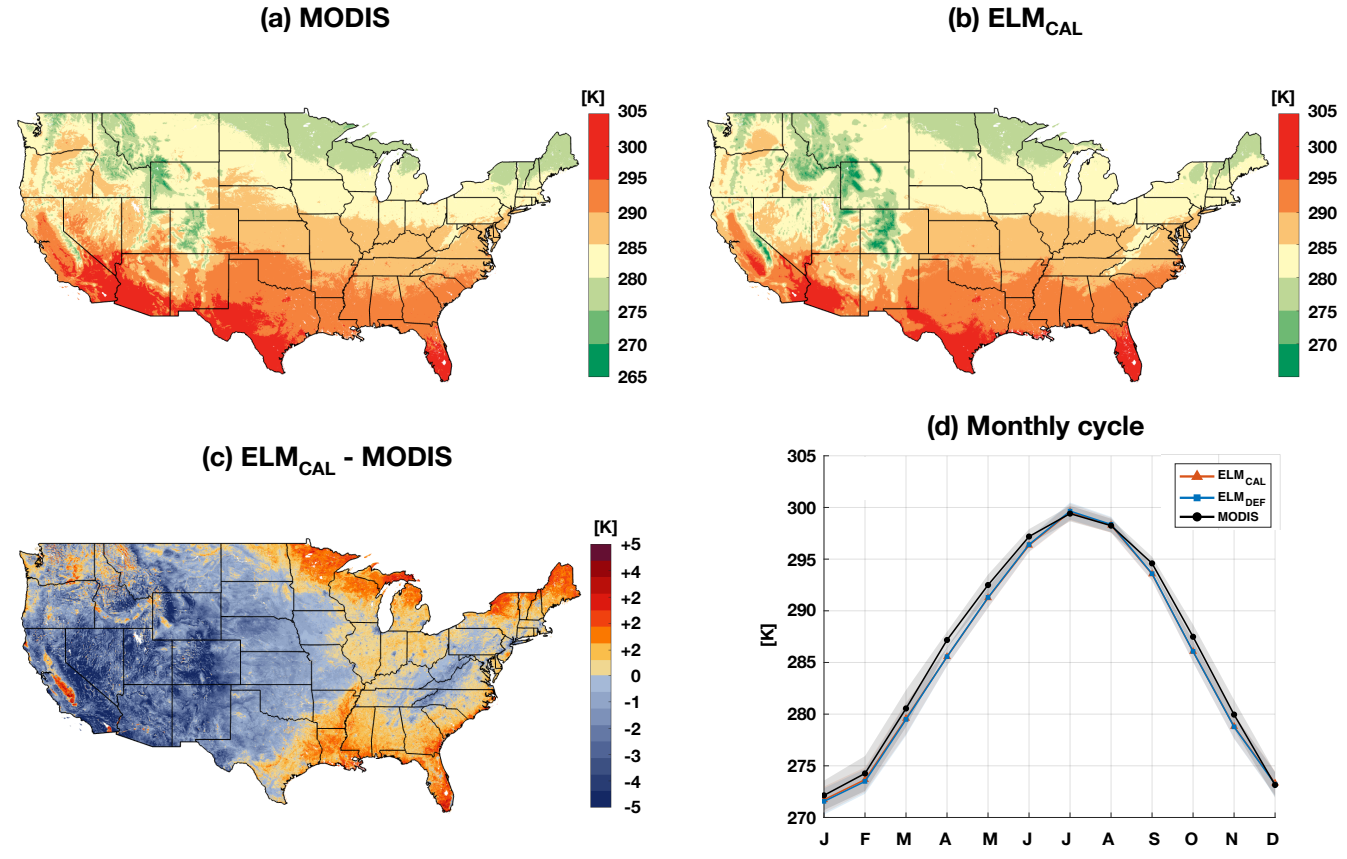


Figure 9. (a) Temporally-averaged MODIS land surface temperature (LST) benchmark, (b) temporally-averaged LST simulated by ELM using calibrated parameters (ELM_{CAL}), (c) temporally-averaged LST bias between ELM_{CAL} and the benchmark, and (d) spatially averaged, monthly time series of LST for ELM_{CAL} and ELM using default parameters, (ELM_{DEF}) against the benchmark. The shading in the time series indicates $\pm$ one spatial standard deviation.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

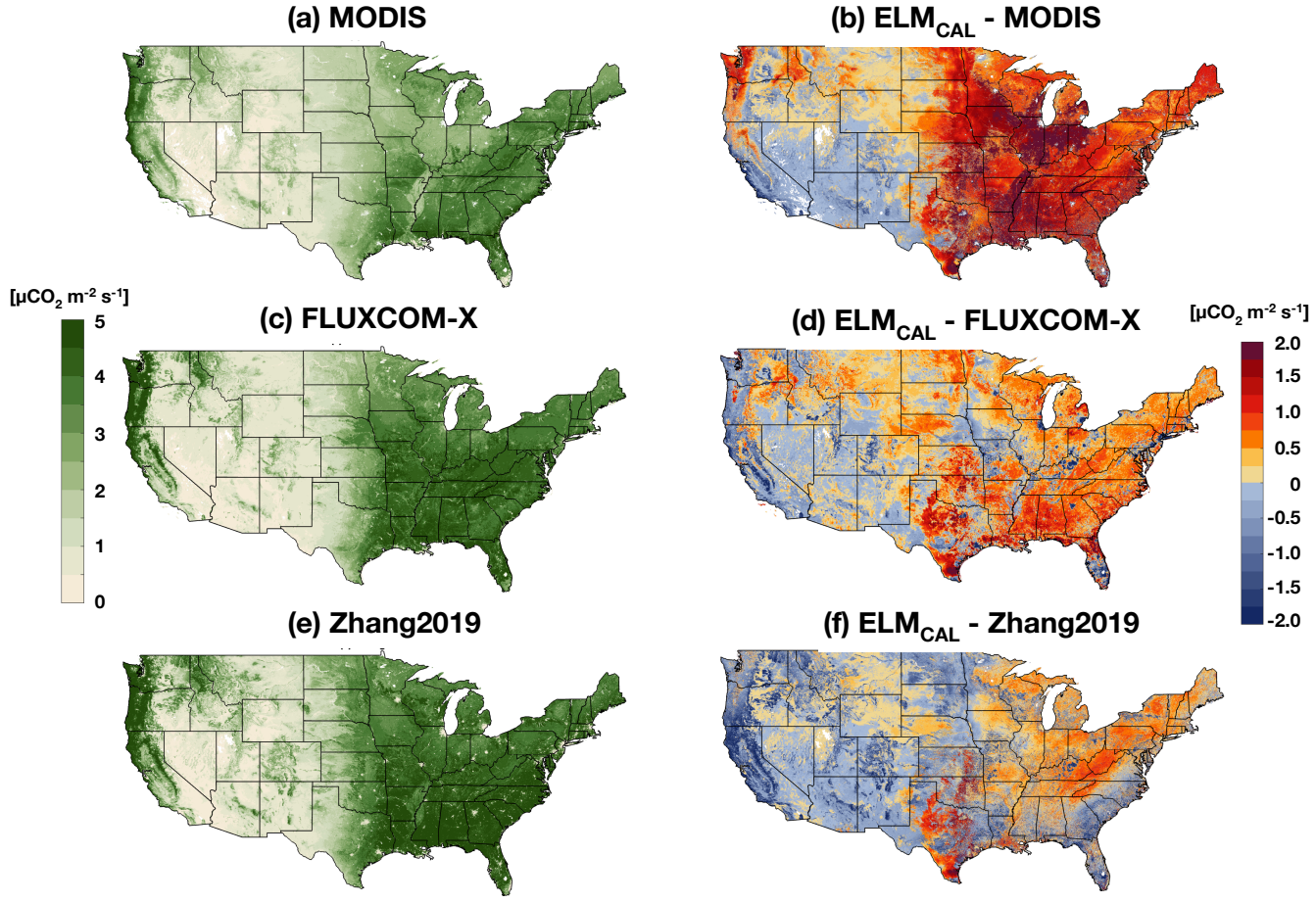


Figure 10. Temporally-averaged GPP benchmark of (a) MODIS, (c) FLUXCOM-X, and (e) Zhang2019, respectively. (b, d, f) Temporally averaged bias between ELM-simulated GPP and the benchmarks. (g) Spatially-averaged, monthly time series of GPP.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

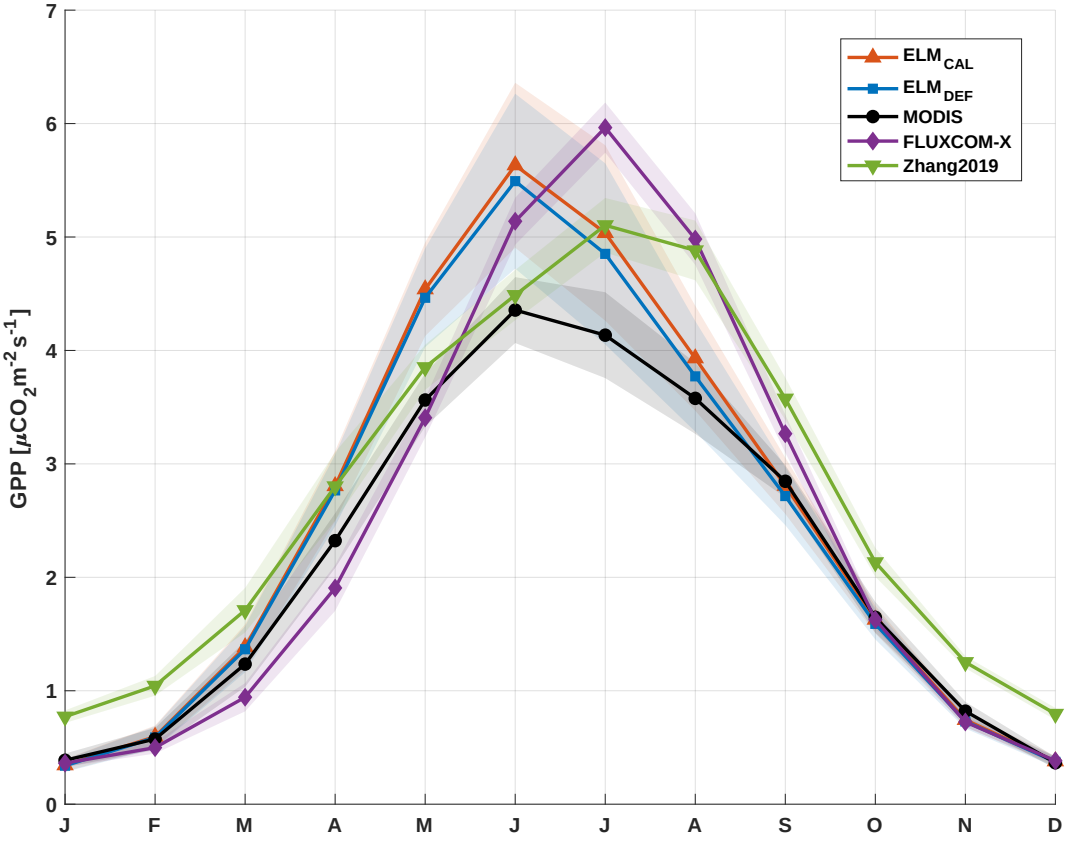


Figure 11. Spatially-averaged, monthly time series of ELM simulation with calibrated parameters, ELM_{CAL}, and default parameters, ELM_{DEF}, against three GPP benchmark datasets.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

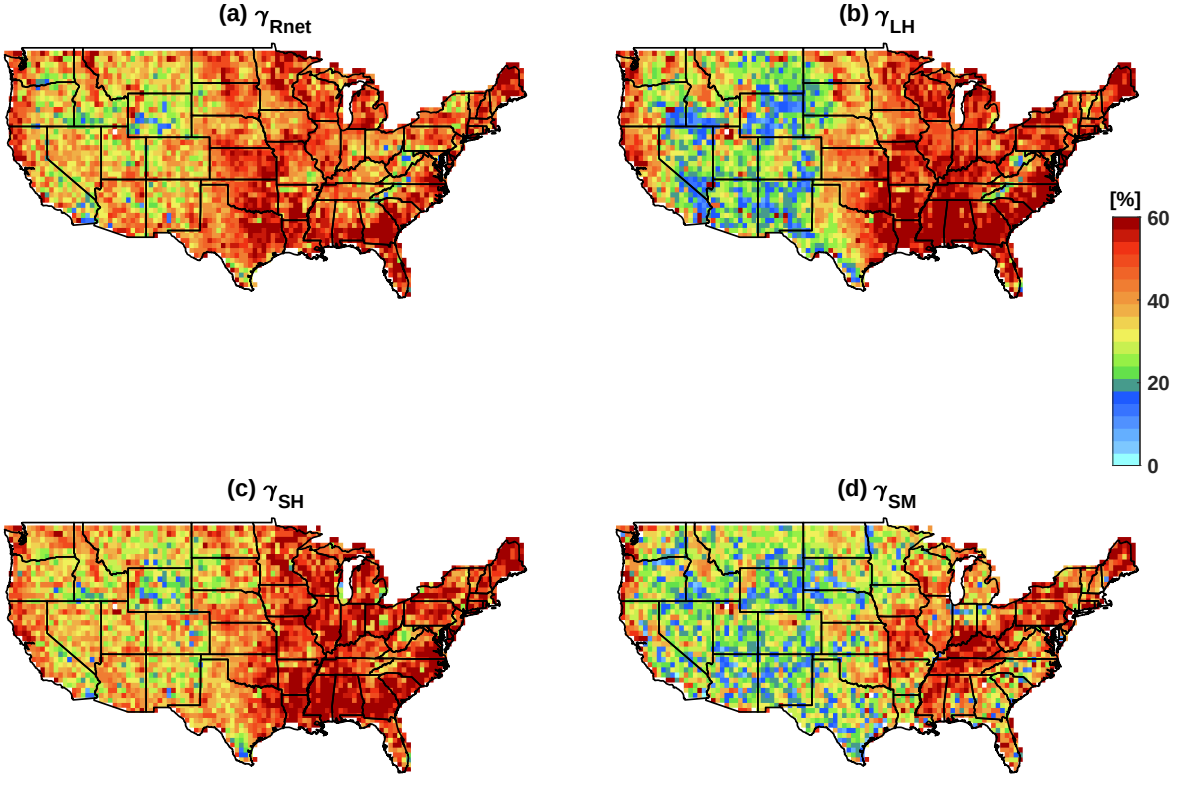


Figure 12. Information loss (γ) of ELM-simulated (a) net radiation (Rnet), (b) latent heat flux (LH), (c) sensible heat flux (SH), and (d) soil moisture (SM).

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

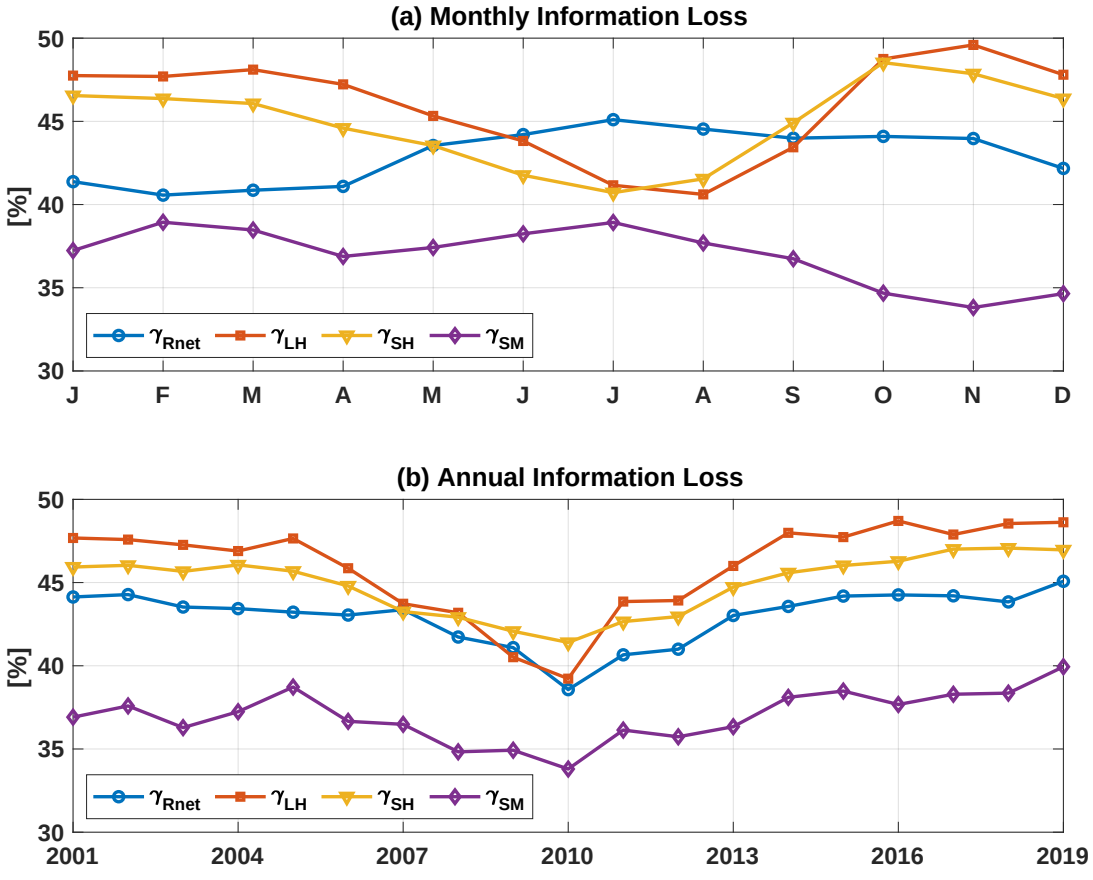


Figure 13. Information loss (γ) of ELM-simulated net radiation (Rnet), latent heat flux (LH), sensible heat flux (SH), and soil moisture (SM) at (a) monthly and (b) annual time scales.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

Table 1. Summary of benchmark datasets used for evaluating ELM simulated water, snow, energy, and carbon dynamics

Variable	Dataset	Time period	Temporal resolution	Spatial resolution	Data Source
WTD	Fan et al. (2013)	N/A	N/A	1 km	http://thredds-gfnl.usc.es/thredds/catalog/GLOBALWTDFTP/catalog.html
f_{ssw}	Global Land Analysis & Discovery	1999 - 2021	Yearly	30m	https://glad.umd.edu/dataset/global-surface-water-dynamics
Q_r	Global Reach-Level Flood Reanalysis	1980-2019	3-hourly	0.05^0	https://www.reachhydro.org/home/records/grfr
Near surface SM	Zheng2023	2000 - 2020	1-day	1 km	https://doi.org/10.11888/RemoteSen.tpdc.272760
SWE	UA	1981 - 2023	1-day	4 km	https://doi.org/10.5067/0GGPB220EX6A
	SNODAS	2003 - present	1-day	1 km	https://doi.org/10.7265/N5TB14TC
	SNOTEL	1979 - 2021	1-day	Point-scale	https://www.nrcs.usda.gov/
LH	MOD16A2GF 6.1	2001 - 2023	8-day	500 m	https://lpdaac.usgs.gov/products/mod16a2v061/
LST	MODIS	2003 - 2020	1-day	1 km	
GPP	Zhang et al. (2019)	2001 - 2023	8-day	500 m	https://developers.google.com/earth-engine/datasets/catalog/CAS_IGSNRR_PML_V2_v018
	FLUXCOM-X	2001 - 2021	Monthly	0.05^0	https://gitlab.gwdg.de/fluxcom/fluxcomxdata
	MOD17A2H	2001 - 2022	8-day	500 m	https://lpdaac.usgs.gov/products/mod17a2hv006/

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

Table 2. Performance metrics (bias, RMSE, and R^2) of the default and calibrated ELM configurations for water table depth (WTD) and the surface water fraction (f_{ssw}). WTD metrics are reported both for all benchmark and for the subset of benchmark with water table depths less than 80 m.

Variable	Default ELM			Calibrated ELM		
	Bias	RMSE	R^2	Bias	RMSE	R^2
WTD [m] using all benchmark	-27.6	54.7	0.27	-7.3	32.8	0.80
WTD [m] using benchmark less than 80 m	-14.5	25.2	0.30	-0.7	2.9	0.99
f_{ssw}	+3.0e-2	5.3e-2	0.14	+1.3e-4	2.0e-2	0.61

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

Table 3. Spatial and temporal evaluation statistics (bias, RMSE, and R^2, with relative bias and rRMSE in parentheses) for calibrated ELM simulated runoff (Q_r), latent heat flux (LH), gross primary productivity (GPP), land surface temperature (LST), and near-surface soil moisture against their respective benchmark datasets. Spatial statistics are computed across grid cells for temporally-averaged fields, whereas temporal statistics are computed across the monthly time series of spatially-averaged values.

Variable	Benchmark	Spatial statistics			Temporal statistics		
		Bias (Relative bias)	RMSE (rRMSE)	R^2	Bias (Relative bias)	RMSE (rRMSE)	R^2
Q_r [mm yr^{-1}]	GRFR	−15.5 (−6.4%)	142.9 (58.8%)	0.92	−25.6 (−10.5%)	101.1 (41.7%)	0.57
LH [W m^{-2}]	MOD17A3H	+7.2 (+19.3%)	11.7 (31.6%)	0.88	+7.9 (+21.3%)	15.0 (40.5%)	0.95
GPP [μCO$_2$m^{-2}s^{-1}]	MOD17A2H	+0.51 (+23.9%)	1.07 (49.9%)	0.86	+0.33 (+15.5%)	0.66 (30.5%)	0.97
	FLUXCOM-X	+0.17 (+7.1%)	0.67 (27.3%)	0.92	+0.05 (+2.1%)	0.71 (29.1%)	0.93
	Zhang2019	−0.21 (−7.3%)	0.84 (29.7%)	0.89	−0.22 (−8.0%)	0.70 (26.1%)	0.94
LST [K]	MODIS	−0.75 (−0.26%)	1.88 (0.66%)	0.96	−0.75 (−0.26%)	0.95 (0.33%)	0.99
Near surface SM	Zheng2023	0.04 (+19.3%)	0.06 (28.8%)	0.83	0.03 (12.4%)	0.03 (13.3%)	0.98
SWE [m]	SNODAS	-6.0e-3 (−46.98%)	2.9e-2 (237.18%)	0.77	-6.0e-3 (−47.19%)	8.0e-3 (66.24%)	0.96
	UA	-6.0e-3 (−48.36%)	2.6e-2 (209.39%)	0.83	-6.0e-3 (−48.37%)	8.9e-3 (70.15%)	0.97
	SNOTEL	−0.12 (−55.5%)	0.21 (115.1%)	0.42	-	-	-
Snow Depth [m]	SNODAS	−0.02 (−44.65%)	0.09 (176.96%)	0.82	−0.02 (−45.35%)	0.03 (64.21%)	0.96
	UA	−0.02 (−37.74%)	0.07 (163.82%)	0.86	−0.02 (−37.74%)	0.02 (57.55%)	0.95

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

Code and data availability. The code used in this study is available at https://zenodo.org/records/22928142 (Tran and Bisht, 2026). The calibrated parameters for ELM are available at https://doi.org/10.5281/zenodo.22945174 (Bisht, 2026), while 1 km parameters for ELM are available at https://doi.org/10.5281/zenodo.10815170 (Li et al., 2024).

Author contributions. **Gautam Bisht**: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing. **Lingcheng Li**: Conceptualization, Methodology, Validation, Investigation, Data Curation, Writing - Review & Editing. **Donghui Xu**: Conceptualization, Methodology, Software, Validation, Investigation, Data Curation, Writing - Review & Editing. **Dalei Hao**: Conceptualization, Methodology, Software, Validation, Investigation, Data Curation, Writing - Review & Editing. **Hoang Tran**: Methodology, Software, Validation, Investigation, Data Curation, Writing - Review & Editing. **L. Ruby Leung**: Conceptualization, Supervision, Funding acquisition, Resources, Project Administration, Writing - Review & Editing.

Competing interests. At least one of the (co-)authors is a member of the editorial board of Geoscientific Model Development.

Acknowledgements. This study is supported by the US Department of Energy (DOE) Office of Science Biological and Environmental Research as part of the Regional and Global Model Analysis (RGMA) program area through the collaborative, multi-program Integrated Coastal Modeling (ICoM) project. This study used DOE's Biological and Environmental Research Earth System Modeling program's Compy computing cluster at Pacific Northwest National Laboratory. Pacific Northwest National Laboratory is operated for the US Department of Energy by Battelle Memorial Institute under contract DE-AC05-76RL01830.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

References

Barrett, A. P.: National operational hydrologic remote sensing center snow data assimilation system (SNODAS) products at NSIDC, https://doi.org/10.7265/N5TB14TC, 2003.

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I., McVicar, T. R., and Adler, R. F.: MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment, Bulletin of the American Meteorological Society, 100, 473–500, https://doi.org/10.1175/BAMS-D-17-0138.1, 2019.

Bisht, G.: Calibrated ELM subsurface drainage and pluvial flood dynamics parameter for CONUS 1km, https://doi.org/10.5281/zenodo.22945174, 2026.

Bisht, G., Huang, M., Zhou, T., Chen, X., Dai, H., Hammond, G. E., Riley, W. J., Downs, J. L., Liu, Y., and Zachara, J. M.: Coupling a three-dimensional subsurface flow and transport model with a land surface model to simulate stream–aquifer–land interactions (CP v1. 0), Geoscientific Model Development, 10, 4539–4562, https://doi.org/10.5194/gmd-10-4539-2017, 2017.

Bisht, G., Riley, W. J., Hammond, G. E., and Lorenzetti, D. M.: Development and evaluation of a variably saturated flow model in the global E3SM Land Model (ELM) version 1.0, Geoscientific Model Development, 11, 4085–4102, https://doi.org/10.5194/gmd-11-4085-2018, 2018.

Bisht, G., Xu, D., Johnson, J., Brown, J., Knepley, M., Adams, M., Feng, D., Hao, D., Engwirda, D., Kumar, M., et al.: Development of a River Dynamical Core for E3SM to simulate compound flooding on Exascale-class heterogeneous supercomputers, Environmental Modelling & Software, p. 106804, https://doi.org/10.1016/j.envsoft.2025.106804, 2025.

Chaney, N. W., Torres-Rojas, L., Vergopolan, N., and Fisher, C. K.: HydroBlocks v0. 2: Enabling a field-scale two-way coupling between the land surface and river networks in Earth system models, Geoscientific Model Development, 14, 6813–6832, https://doi.org/10.5194/gmd-14-6813-2021, 2021.

Cheng, Y., Zhao, L., Chakraborty, T., Oleson, K., Demuzere, M., Liu, X., Che, Y., Liao, W., Zhou, Y., and Li, X.: U-Surf: a global 1 km spatially continuous urban surface property dataset for kilometer-scale urban-resolving Earth system modeling, Earth System Science Data, 17, 2147–2174, https://doi.org/10.5194/essd-17-2147-2025, 2025.

Collier, N., Hoffman, F. M., Lawrence, D. M., Keppel-Aleks, G., Koven, C. D., Riley, W. J., Mu, M., and Randerson, J. T.: The International Land Model Benchmarking (ILAMB) system: design, theory, and implementation, Journal of Advances in Modeling Earth Systems, 10, 2731–2754, https://doi.org/10.1029/2018MS001354, 2018.

Dickinson, R. E.: Land surface processes and climate—Surface albedos and energy balance, in: Advances in geophysics, vol. 25, pp. 305–353, Elsevier, https://doi.org/10.1016/S0065-2687(08)60176-4, 1983.

Donahue, A. S., Caldwell, P. M., Bertagna, L., Beydoun, H., Bogenschutz, P. A., Bradley, A., Clevenger, T. C., Foucar, J., Golaz, C., Guba, O., et al.: To exascale and beyond—The Simple Cloud-Resolving E3SM Atmosphere Model (SCREAM), a performance portable global atmosphere model for cloud-resolving scales, Journal of Advances in Modeling Earth Systems, 16, e2024MS004 314, https://doi.org/10.1029/2024MS004314, 2024.

Fan, Y., Li, H., and Miguez-Macho, G.: Global patterns of groundwater table depth, Science, 339, 940–943, https://doi.org/10.1126/science.1229881, 2013.

Fang, Y., Tran, H. V., and Leung, L. R.: Subsurface hydrological controls on the short-term effects of hurricanes on nitrate–nitrogen runoff loading: a case study of Hurricane Ida using the Energy Exascale Earth System Model (E3SM) Land Model (v2. 1), Geoscientific Model Development, 18, 19–32, https://doi.org/10.5194/gmd-18-19-2025, 2025.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

Farmani, M. A., Behrangi, A., Gupta, A., Tavakoly, A., Geheran, M., and Niu, G.-Y.: Do land models miss key soil hydrological processes controlling soil moisture memory?, Hydrology and Earth System Sciences, 29, 547–566, https://doi.org/10.5194/hess-29-547-2025, 2025a.

Farmani, M. A., Tavakoly, A., Behrangi, A., Qiu, Y., Gupta, A., Jawad, M., Sohi, H. Y., Zhang, X., Geheran, M., and Niu, G.-Y.: Improving streamflow predictions in the arid Southwestern United States through understanding of baseflow generation mechanisms, Water Resources Research, 61, e2024WR039 479, https://doi.org/10.1029/2024WR039479, 2025b.

Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., Abeshu, G., Anantharaj, V., Asay-Davis, X. S., Bader, D. C., et al.: The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution, Journal of Advances in Modeling Earth Systems, 11, 2089–2129, https://doi.org/10.1029/2018MS001603, 2019.

Golaz, J.-C., Van Roekel, L. P., Zheng, X., Roberts, A. F., Wolfe, J. D., Lin, W., Bradley, A. M., Tang, Q., Maltrud, M. E., Forsyth, R. M., et al.: The DOE E3SM model version 2: Overview of the physical model and initial model evaluation, Journal of Advances in Modeling Earth Systems, 14, e2022MS003 156, https://doi.org/10.1029/2022MS003156, 2022.

Hao, D., Bisht, G., Gu, Y., Lee, W.-L., Liou, K.-N., and Leung, L. R.: A parameterization of sub-grid topographical effects on solar radiation in the E3SM Land Model (version 1.0): implementation and evaluation over the Tibetan Plateau, Geoscientific Model Development, 14, 6273–6289, https://doi.org/10.5194/gmd-14-6273-2021, 2021.

Hao, D., Bisht, G., Huang, M., Ma, P.-L., Tesfa, T., Lee, W.-L., Gu, Y., and Leung, L. R.: Impacts of sub-grid topographic representations on surface energy balance and boundary conditions in the E3SM land model: A case study in Sierra Nevada, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 862, https://doi.org/10.1029/2021MS002862, 2022.

Hao, D., Bisht, G., Rittger, K., Stillinger, T., Bair, E., Gu, Y., and Leung, L. R.: Evaluation of E3SM land model snow simulations over the western United States, The Cryosphere, 17, 673–697, https://doi.org/10.5194/tc-17-673-2023, 2023.

Hao, D., Bisht, G., Li, L., and Leung, L. R.: Representing fine-scale topographic effects on surface radiation balance in hyper-resolution land surface models, Journal of Advances in Modeling Earth Systems, 17, e2025MS004 987, https://doi.org/10.1029/2025MS004987, 2025.

Hazenberg, P., Fang, Y., Broxton, P., Gochis, D., Niu, G.-Y., Pelletier, J., Troch, P., and Zeng, X.: A hybrid-3D hillslope hydrological model for use in E arth system models, Water Resources Research, 51, 8218–8239, https://doi.org/10.1002/2014WR016842, 2015.

Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Alonso-Balmaseda, M., Balsamo, G., Bechtolg, P., et al.: Operational global reanalysis: progress, future directions and synergies with NWP, ECMWF Newsletter, https://doi.org/10.21957/tkic6g3wm, 2018.

Jackson, T. L., Feddema, J. J., Oleson, K. W., Bonan, G. B., and Bauer, J. T.: Parameterization of Urban Characteristics for Global Climate Modeling, Annals of the Association of American Geographers, 100, 848–865, https://doi.org/10.1080/00045608.2010.497328, 2010.

Keetz, L. T., Aalstad, K., Fisher, R., Poppe Terán, C., Naz, B., Pirk, N., Yilmaz, Y., and Skarpaas, O.: Inferring parameters in a complex land surface model by combining data assimilation and machine learning, Journal of Advances in Modeling Earth Systems, 17, e2024MS004 542, https://doi.org/10.1029/2024MS004542, 2025.

Kollet, S., Belleflamme, A., Condon, L., Fahad, M., Goergen, K., Maxwell, R., and Naz, B.: Global groundwater modeling: Proof-of-concept of 3D variably saturated flow simulation at kilometer resolution, Journal of Hydrology X, p. 100213, https://doi.org/10.1016/j.hydroa.2025.100213, 2025.

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., Van Kampenhout, L., Kennedy, D., et al.: The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty, Journal of Advances in Modeling Earth Systems, 11, 4245–4287, https://doi.org/10.1029/2018MS001583, 2019.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

Li, L., Bisht, G., Hao, D., and Leung, L. R.: Global 1 km land surface parameters for kilometer-scale Earth system modeling, Earth System Science Data, 16, 2007–2032, https://doi.org/10.5194/essd-16-2007-2024, 2024.

Li, L., Bisht, G., Xu, D., Hao, D., Liao, C., Feng, D., Tan, Z., Li, H., Shi, M., Zhou, T., et al.: Systematic evaluation of atmospheric forcing, surface datasets, and mesh effects on Kilometer-Scale land surface and river modeling, Journal of Hydrology, p. 135500, https://doi.org/10.1016/j.jhydrol.2026.135500, 2026.

Lv, M., Yang, Z.-L., Xu, Z., Dan, L., Lv, M., and Zheng, H.: A Soil Moisture-Dependent Model to Simulate Water Table Depth and Proportions of Surface and Subsurface Runoff and Its Validation at the Basin Scale, Journal of Geophysical Research: Atmospheres, 126, https://doi.org/10.1029/2020JD033661, 2021.

Ma, Y., He, T., Aguilar, C., Pimentel, R., Liang, S., McVicar, T. R., Hao, D., Xiao, X., and Liu, X.: Evaluating Topographic Effects on Kilometer-Scale Satellite Downward Shortwave Radiation Products: A Case Study in Mid-Latitude Mountains, IEEE Transactions on Geoscience and Remote Sensing, 62, 1–16, https://doi.org/10.1109/TGRS.2024.3365865, 2024.

Ma, Y., Condon, L. E., Koch, J., Bennett, A., Defnet, A., Tijerina-Kreuzer, D., Melchior, P., and Maxwell, R. M.: High resolution US water table depth estimates reveal quantity of accessible groundwater, Communications Earth & Environment, 7, 45, https://doi.org/10.1038/s43247-025-03094-3, 2026.

Maxwell, R. M.: A terrain-following grid transform and preconditioner for parallel, large-scale, integrated hydrologic modeling, Advances in Water Resources, 53, 109–117, https://doi.org/10.1016/j.advwatres.2012.10.001, 2013.

Maxwell, R. M. and Condon, L. E.: Connections between groundwater flow and transpiration partitioning, Science, 353, 377–380, https://doi.org/10.1126/science.aaf7891, 2016.

Miura, Y. and Yoshimura, K.: Development and verification of a three-dimensional variably saturated flow model for assessment of future global water resources, Journal of Advances in Modeling Earth Systems, 12, e2020MS002 093, https://doi.org/10.1029/2020MS002093, 2020.

Mu, Q., Zhao, M., and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote sensing of environment, 115, 1781–1800, https://doi.org/10.1016/j.rse.2011.02.019, 2011.

Nelson, J. A., Walther, S., Gans, F., Kraft, B., Weber, U., Novick, K., Buchmann, N., Migliavacca, M., Wohlfahrt, G., Šigut, L., et al.: X-BASE: the first terrestrial carbon and water flux products from an extended data-driven scaling framework, FLUXCOM-X, Biogeosciences, 21, 5079–5115, https://doi.org/10.5194/bg-21-5079-2024, 2024.

Oleson, K. W., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S., Li, F., Riley, W. J., Subin, Z. M., et al.: Technical description of version 4.0 of the Community Land Model (CLM), NCAR Tech. Note NCAR/TN-478+ STR, 257, 1–257, https://doi.org/10.5065/D6FB50WZ, 2010.

Pickens, A. H., Hansen, M. C., Hancher, M., Stehman, S. V., Tyukavina, A., Potapov, P., Marroquin, B., and Sherani, Z.: Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series, Remote Sensing of Environment, 243, 111 792, https://doi.org/10.1016/j.rse.2020.111792, 2020.

Qiu, H., Bisht, G., Li, L., Hao, D., and Xu, D.: Development of inter-grid-cell lateral unsaturated and saturated flow model in the E3SM Land Model (v2. 0), Geoscientific Model Development, 17, 143–167, https://doi.org/10.5194/gmd-17-143-2024, 2024.

Rasmussen, R. M., Chen, F., Liu, C., Ikeda, K., Prein, A., Kim, J.-h., Schneider, T., Dai, A., Gochis, D., Dugger, A., et al.: CONUS404: The NCAR–USGS 4-km long-term regional hydroclimate reanalysis over the CONUS, Bulletin of the American Meteorological Society, 104, E1382–E1408, https://doi.org/10.1175/BAMS-D-21-0326.1, 2023.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.

Running, S., Mu, Q., and Zhao, M.: Mod16a2 modis/terra net evapotranspiration 8-day l4 global 500m sin grid v006, (No Title), https://doi.org/10.5067/MODIS/MOD16A2.006, 2017.

Silvestri, S., Wagner, G. L., Constantinou, N. C., Hill, C. N., Campin, J.-M., Souza, A. N., Bishnu, S., Churavy, V., Marshall, J., and Ferrari, R.: A GPU-based ocean dynamical core for routine mesoscale-resolving climate simulations, Journal of Advances in Modeling Earth Systems, 17, e2024MS004 465, https://doi.org/10.1029/2024MS004465, 2025.

Swenson, S. C., Clark, M., Fan, Y., Lawrence, D. M., and Perket, J.: Representing intrahillslope lateral subsurface flow in the community land model, Journal of Advances in Modeling Earth Systems, 11, 4044–4065, https://doi.org/10.1029/2019MS001833, 2019.

Tang, Q., Golaz, J.-C., Van Roekel, L. P., Taylor, M. A., Lin, W., Hillman, B. R., Ullrich, P. A., Bradley, A. M., Guba, O., Wolfe, J. D., et al.: The fully coupled regionally refined model of E3SM version 2: Overview of the atmosphere, land, and river, Geoscientific Model Development, 16, 3953–3995, https://doi.org/10.5194/gmd-16-3953-2023, 2023.

Tran, H. and Bisht, G.: E3SM version used for CONUS 1km study, https://doi.org/10.5281/zenodo.22928142, 2026.

Vergopolan, N., Sheffield, J., Chaney, N. W., Pan, M., Beck, H. E., Ferguson, C. R., Torres-Rojas, L., Eigenbrod, F., Crow, W., and Wood, E. F.: High-resolution soil moisture data reveal complex multi-scale spatial variability across the United States, Geophysical Research Letters, 49, e2022GL098 586, https://doi.org/10.1029/2022GL098586, 2022.

Wang, D., Wang, C., Cao, Q., Krishna, J., Zheng, W., Schwartz, P., Yuan, F., Ricciuto, D., Thornton, P., and Kao, S.-C.: Enabling kilometer-scale E3SM land model simulation over North America: A new integrated framework solution, The International Journal of High Performance Computing Applications, p. 10943420261446003, https://doi.org/10.1177/10943420261446003, 2026.

Wood, E. F., Roundy, J. K., Troy, T. J., Van Beek, L., Bierkens, M. F., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., et al.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resources Research, 47, https://doi.org/10.1029/2010WR010090, 2011.

Xie, Z., Wang, L., Wang, Y., Liu, B., Li, R., Xie, J., Zeng, Y., Liu, S., Gao, J., Chen, S., et al.: Land surface model CAS-LSM: Model description and evaluation, Journal of Advances in Modeling Earth Systems, 12, e2020MS002 339, https://doi.org/10.1029/2020MS002339, 2020.

Xu, D., Bisht, G., Sargsyan, K., Liao, C., and Leung, L. R.: Using a surrogate-assisted Bayesian framework to calibrate the runoff-generation scheme in the Energy Exascale Earth System Model (E3SM) v1, Geoscientific Model Development, 15, 5021–5043, https://doi.org/10.5194/gmd-15-5021-2022, 2022.

Xu, D., Bisht, G., Tan, Z., Liao, C., Zhou, T., Li, H.-Y., and Leung, L. R.: Disentangling the hydrological and hydraulic controls on streamflow variability in Energy Exascale Earth System Model (E3SM) V2–a case study in the Pantanal region, Geoscientific Model Development, 17, 1197–1215, https://doi.org/10.5194/gmd-17-1197-2024, 2024a.

Xu, D., Bisht, G., Tan, Z., Sinha, E., Di Vittorio, A. V., Zhou, T., Ivanov, V. Y., and Leung, L. R.: Climate change will reduce North American inland wetland areas and disrupt their seasonal regimes, Nature Communications, 15, 2438, https://doi.org/10.1038/s41467-024-45286-z, 2024b.

Yang, Y., Pan, M., Lin, P., Beck, H. E., Zeng, Z., Yamazaki, D., David, C. H., Lu, H., Yang, K., Hong, Y., et al.: Global reach-level 3-hourly river flood reanalysis (1980–2019), Bulletin of the American Meteorological Society, 102, E2086–E2105, https://doi.org/10.1175/BAMS-D-20-0057.1, 2021.

Zeng, X., Broxton, P., and Dawson, N.: Snowpack change from 1982 to 2016 over conterminous United States, Geophysical Research Letters, 45, 12–940, https://doi.org/10.1029/2018GL079621, 2018.

https://doi.org/10.5194/egusphere-2026-5918
Preprint. Discussion started: 26 October 2026
© Author(s) 2026. CC BY 4.0 License.



Zhang, T., Zhou, Y., Zhu, Z., Li, X., and Asrar, G. R.: A global seamless 1 km resolution daily land surface temperature dataset (2003–2020), Earth System Science Data, 14, 651–664, https://doi.org/10.5194/essd-14-651-2022, 2022.

Zhang, Y., Kong, D., Gan, R., Chiew, F. H., McVicar, T. R., Zhang, Q., and Yang, Y.: Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017, Remote sensing of environment, 222, 165–182, https://doi.org/10.1016/j.rse.2018.12.031, 2019.

Zheng, C., Jia, L., and Zhao, T.: A 21-year dataset (2000–2020) of gap-free global daily surface soil moisture at 1-km grid resolution, Scientific Data, 10, 139, https://doi.org/10.1038/s41597-023-01991-w, 2023.