U.S. DEPARTMENT OF ENERGY | Office of Science

DOE/SC-CM-24-001

FY 2024 First Quarter Performance Metric: Developing and demonstrating a Kilometer-Scale Land Simulation Capability for Modeling the Terrestrial Water Cycle over the Contiguous U.S.

Lai-Yung Ruby Leung
Gautam Bisht
Lingcheng Li

January 2024

DISCLAIMER

Contents

Figures

1.0 Product Definition

Kilometer-scale (k-scale) modeling allows for explicit modeling of physical processes that are poorly represented in climate models with typical resolutions of 25-100 km, thus providing opportunities to significantly improve the accuracy of climate simulations (Slingo et al. 2022). The recent advancements in computing power are making k-scale regional and global simulations using Land Surface Models (LSMs) and Earth System Models (ESMs) increasingly feasible (Condon et al. 2020, Caldwell et al. 2021). Groundwater is a vital human water resource that provides 20-30% of global freshwater withdrawals (Döll 2009). Previous k-scale LSM studies for specific watersheds or basins have demonstrated the impacts of fine scale-structures on terrestrial hydrologic processes including groundwater dynamics (Maxwell and Kollet 2008, Fan et al. 2013).

However, the parameters of LSMs within ESMs being run at the k scale are typically derived from coarse-resolution data sets or outdated data sets. Consequently, k-scale modeling may not accurately represent fine-scale land surface heterogeneity unless high-resolution land surface parameters at the kilometer or finer scales are used. Additionally, LSMs will have to be recalibrated at k scale to accurately simulate terrestrial processes (Ruiz-Vásquez et al. 2023) and the recalibration is expected to have a significantly large computational cost at such high spatial resolution.

In this report, we develop a first-of-a-kind k-scale global land simulation capability and demonstrate this for modeling the terrestrial water cycle over the contiguous U.S. in ESMs. To enable this capability, we first develop a new set of global land surface parameters at 1-km resolution by using the newest high-resolution data sources for multiple years by combining the latest and most accurate available global data sets to provide input data for ESMs, including the U.S. Department of Energy's flagship Energy Exascale Earth System Model (E3SM) (Leung et al. 2020). Second, we produce an initial 5-year simulation using E3SM Land Model version 2 (ELMv2) over the contiguous United States (CONUS) at 1-km resolution using the newly developed land surface parameters. A spatial scaling analysis was performed to underscore the value of the high-resolution land surface parameters, and AI/ML methods were used to identify the most important land surface parameters and climate conditions that drive the spatial variability at k scale and hence, spatial information loss as resolutions were coarsened in ELMv2 simulations. Using the newly developed 1-km input data, the k-scale simulations show significant spatial heterogeneity of soil moisture, latent heat, emitted longwave radiation, and absorbed shortwave radiation, which is expected to have important effect on modeling of land-atmosphere interactions. Third, we demonstrate significant improvements in the ELM-simulated water table depth over CONUS by calibrating ELMv2's subsurface drainage parameterization for k-scale modeling.

2.0 Product Documentation

This report documents a k-scale land simulation capability, which includes the development of global k-scale land surface parameters for LSMs (2.1), the application of the newly developed surface parameters in k-scale ELMv2 simulation over CONUS and analysis that demonstrates the importance of the k-scale surface parameters data set (2.2), and the calibration of ELMv2's subsurface hydrologic processes to improve the prediction of groundwater at a 1-km resolution over CONUS (2.3).

2.1 Global 1-km Land Surface Parameters

The development of global 1-km land surface parameters takes advantage of global high-resolution satellite data that have become available in the last two decades. New parameters in four categories were developed: (1) Land Use Land Cover (LULC)-related parameters (e.g., spatial distribution of vegetation, lakes, urban areas, and glaciers), (2) vegetation-related parameters (e.g., Leaf Area Index (LAI) and Stem Area Index(SAI)), (3) soil-related parameters (e.g., soil texture and soil organic matter), and (4) topography-related parameters (e.g., slope, aspect, and sub-grid topographic factors). These parameters support global land surface modeling of the biophysical, hydrological, and biogeochemical processes at 1-km resolution.

2.2 ELM 1-km CONUS Simulation and Spatial Scaling Analysis

To demonstrate a new k-scale land simulation capability, ELMv2 simulations were conducted over CONUS at the resolution of 1 km, using the newly developed 1-km land surface parameters for 2010, which is representative of present-day conditions. To isolate the impact of the high-resolution land surface parameters, the simulation was driven by atmospheric forcing from the Global Soil Wetness Project Phase 3 (GSWP3; Kim 2017) with a coarse spatial resolution of 0.5°. The simulations were performed for approximately 12 million grids over five years (2010–2014), and the last year of the simulation was used for analysis.

The method described in Vergopolan (2022) was used to perform a spatial scaling analysis on the 1-km ELMv2 simulation to quantify the impact of the spatial heterogeneity of land surface parameters on the simulation. This analysis focuses on four ELM variables representing key land surface processes for the energy and water cycles: surface layer soil moisture (SM), latent heat flux (LH), emitted longwave radiation (ELR), and absorbed shortwave radiation (ASR). By upscaling the land surface parameters and model simulations from 1 km (=1/120°) to coarser spatial scales, λ_{scale}, of 1/60°, 1/40°, 1/30°, 1/24°, 1/20°, and 1/10°, we calculated the spatial standard deviation (σ_{scale}) within each 0.5° × 0.5° box for each spatial scale. This allows us to quantify the changes in spatial variability from the original 1-km resolution to the different coarser spatial scales, $\sigma_{scale}/\sigma_{1\,km}$, and the information loss at coarser resolutions, $\gamma = (1 - \sigma_{scale}/\sigma_{1\,km}) \times 100\%$.

XML methods were then used to assess the influence of land surface parameters on the spatial variability and information loss. For this, we trained a machine learning model to predict the spatial variability of each simulated variable (SM, LH, ELR, ASR) within each 0.5° × 0.5° box using the spatial variability and mean of the land surface parameters and the spatial mean of precipitation and temperature as predictor variables. After training the machine learning model, we quantified the relative importance of the predictors using a game theoretic approach. We showed that among the various land surface parameters, the spatial heterogeneity of soil properties, vegetation properties, and topography are more important in driving the spatial variability of the simulations and the information loss due to coarsening of the spatial resolution.

2.3 Improving k-Scale Simulation of Water Table Depth

The performance of ELMv2-simulated groundwater dynamics at k-scale is expected to be poor if the default, coarse-scale hydrologic model parameters are used without any model recalibration. Several

previous studies (Bisht et al. 2018, Xu et al. 2023) have found that a subsurface runoff generation parameter, f_{drain}, is the most sensitive parameter determining the groundwater dynamics in ELMv2. For the 1-km CONUS simulation described in Section 2.2, the default f_{drain} value of 2.5 m^{-1} was used. To improve the simulation, the f_{drain} parameter was calibrated using the global 1-km water table depth (WTD) data set of Fan et al. (2013) (hereafter referred to as F2013) as the benchmark. Following the calibration procedure described in Bisht et al. (2018), seven ELMv2 simulations were performed with spatially homogeneous f_{drain} values of 0.2, 0.3, 0.5, 0.5, 1.0, 2.5, and 5.0 m^{-1}. By developing a nonlinear functional relationship between f_{drain} and WTD for each grid cell within the domain, an optimal f_{drain} value was estimated for each grid cell that reduces the difference between the simulated and observed WTD. Using the optimal f_{drain} value for each grid cell, a new 1-km CONUS simulation was performed and evaluated against the F2013 data set. The new simulation shows significant improvements in simulating water table depth compared to the simulation with the default and spatially uniform f_{drain} value.

3.0 Results

3.1 Global 1-km Land Surface Parameters

To facilitate the high-resolution ELM land surface simulations described in section 2.2, a new set of global land surface parameters at 1 km is needed. A description of the methodology used to develop the land surface parameters can be found in Li et al. (2023). Table 1 summarizes the spatiotemporal resolution and the data sources of the surface parameters processed to develop the global 1-km (i.e., 1/120°) land surface parameters for LSMs. Global LAI at 1-km resolution generally shows high values in humid and warm regions, such as tropical rainforests, southeastern U.S., and southern Asia, and low values over arid or cold regions, such as central Australia, southwestern U.S., Middle East, Central Asia, and northern Canada (Figure 1a). At high resolution, the LAI data set clearly reflects the detailed heterogeneity of vegetation distributions. In subregion R1 (Figure 1b), a relatively small LAI is distributed over mountain ridges and zero LAI values over water surfaces (e.g., lakes). In subregion R2 (Figure 1c), the LAI pattern shows a large proportion of forest fragmentation caused by deforestation. In subregion R3 (Figure 1d), the LAI shows the distribution of agricultural land along with the river, river mouth, and lakes under an arid climate. R4 shows how urbanization affects vegetation distributions (Figure 1e). The distribution of PFTs and other non-vegetation land units for 2010 is shown in Figure 2. High-resolution LULC types over multiple years can benefit studies related to LULC changes like urbanization and deforestation.

Table 1. The spatiotemporal resolution and data source for generating 1-k global land surface parameters.

Category	Land surface parameters	Description
LULC	Plant Function Types, Lake, Glacier, Urban	• Resolution: 1 km, yearly, 2001-2020 • Data source: 500 m, yearly, MODIS collection 6 (Friedl et al., 2022)
Vegetation	Leaf Area Index, Stem Area Index	• Resolution: 1 km, monthly, 2001-2020 • Data source: 450 m, 8-day, reprocessed MODIS collection 6 LAI (Yuan et al., 2011; Friedl et al., 2022)
	Canopy top height, Canopy bottom height	• Resolution: 1 km, temporally static • Data source: 10 m, vegetation canopy height (Lang et al., 2023)
Soil	Percent sand, Percent clay	• Resolution: 1 km, temporally static • Data source: 250 m, Soilgrid v2 (Poggio et al., 2021)
	Soil organic matter	
Topography	Elevation, Slope	• Resolution: 1 km, temporally static • Data source: 90 m, MERIT Hydro elevation (Yamazaki et al., 2019)
	Standard deviation of elevation, aspect, sky view factor, and terrain view factor	• Resolution: 1 km, temporally static • Data source: 90 m, Hydro elevation (Yamazaki et al., 2019)

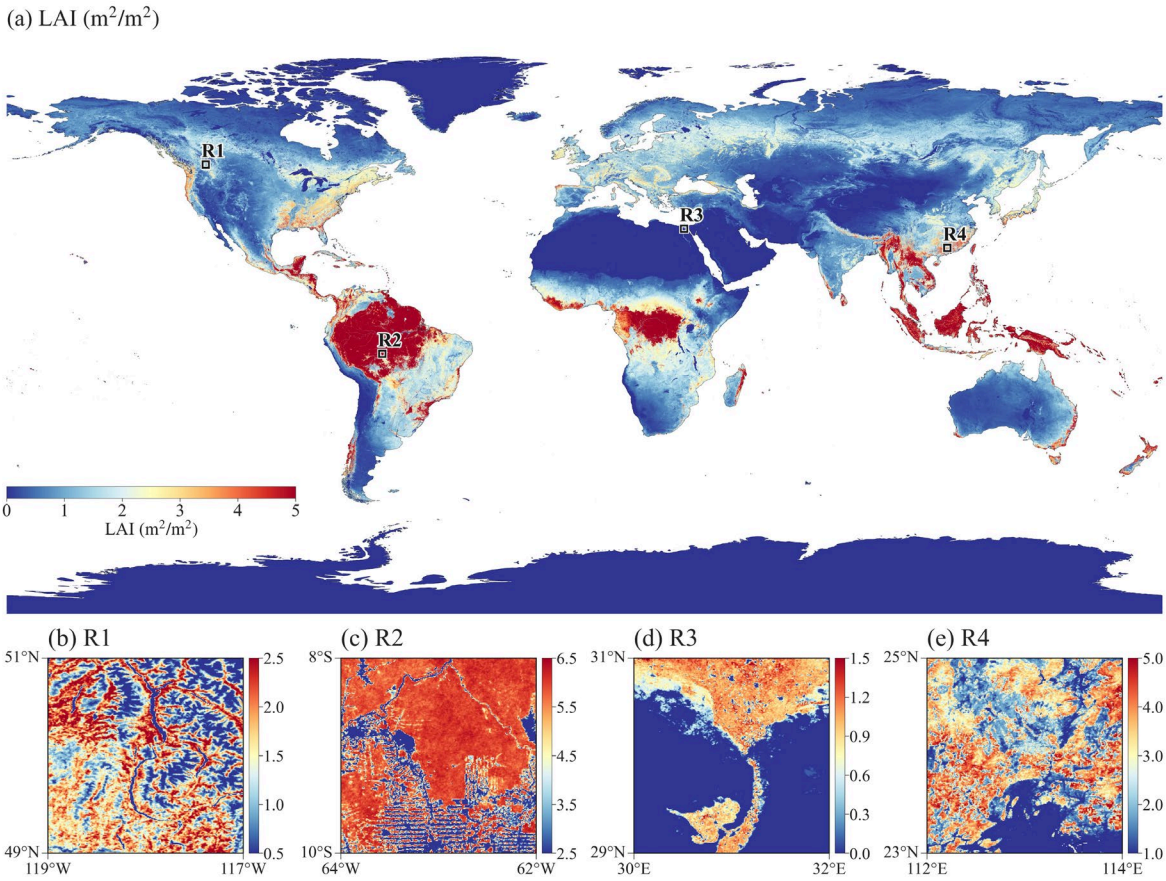

Figure 1. The spatial pattern of annual average Leaf Area Index (LAI) for 2010 over (a) global land and (b)-(e) four subregions R1 to R4 within 2-degree boxes marked in (a). Large spatial heterogeneity driven by different factors is shown in the subregions.

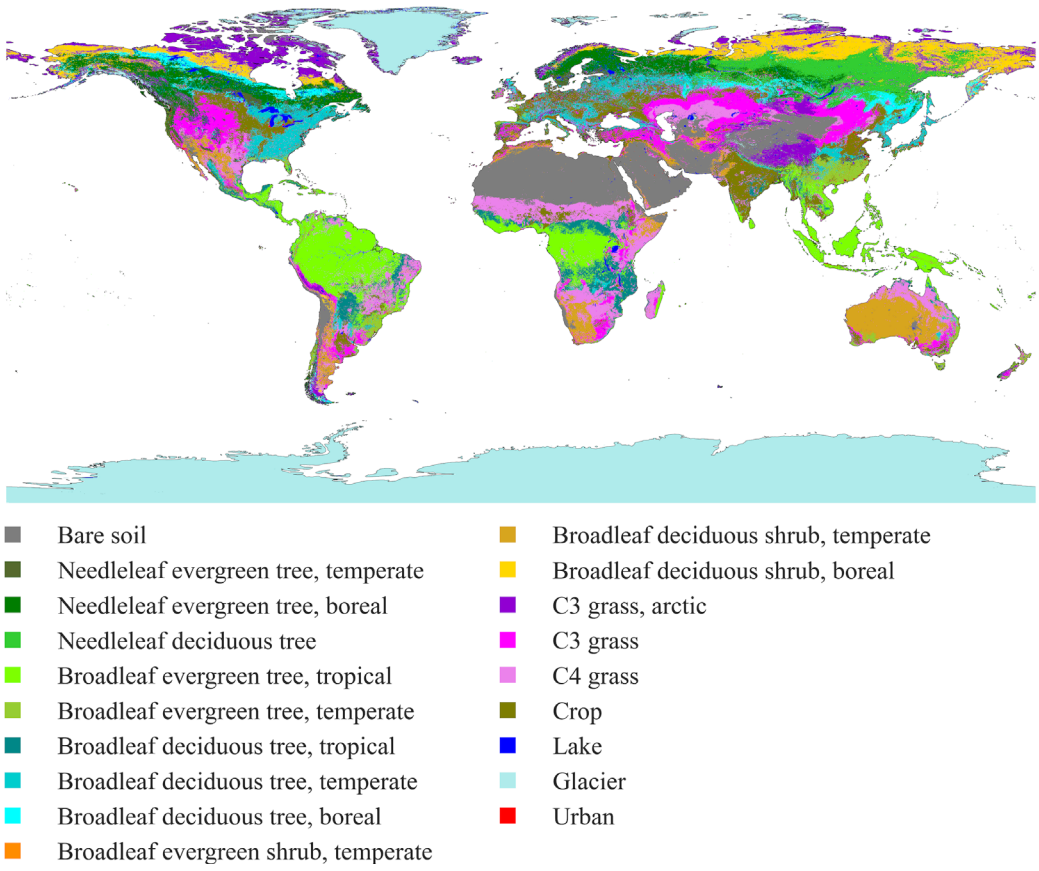

Figure 2. Global Land Use Land Cover (LULC) distribution in year 2010.

3.2 Spatial Scaling Analysis of Soil Moisture and Energy Fluxes in the 1-km ELM CONUS Simulation

To demonstrate the k-scale land simulation capability of the E3SM land model, ELMv2 simulations at 1-km resolution were performed using the new 1-km input data described in section 3.1. Surface water and energy fluxes exemplified by soil moisture (SM), latent heat (LH), emitted longwave radiation (ELR), and absorbed shortwave radiation (ASR) display significant spatial heterogeneity over CONUS (Figure 3) and follow approximately normal distributions, with average values of 0.3 m^3/m^3, 39.0 W/m^2, 371.7 W/m^2, 156.7 W/m^2, respectively. SM shows drier conditions over the West and Southwest and wetter conditions over the Midwest, Corn Belt, Mississippi River basin, and Northeast (Figure 3a).

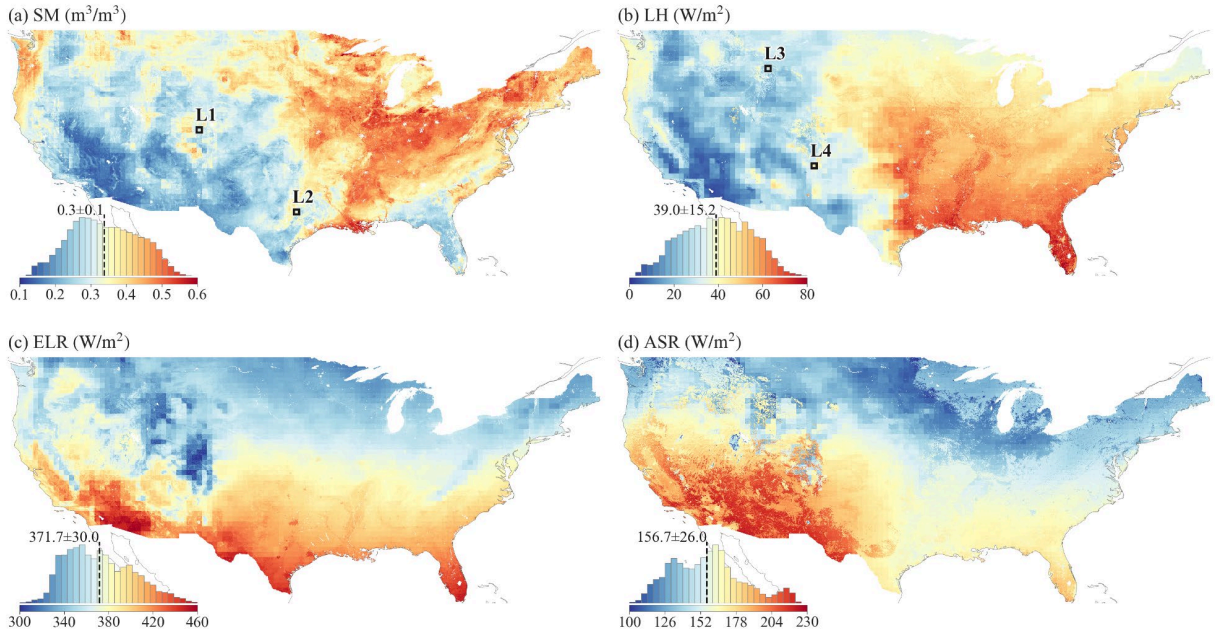

Figure 3. The annual mean of 1-km simulations of (a) soil moisture (SM), (b) latent heat flux (LH), (c) emitted longwave radiation (ELR), and (d) absorbed shortwave radiation (ASR) over CONUS. The 0.5° × 0.5° boxes marked as L1, L2, L3, and L4 in (a) and (b) are selected to demonstrate the spatial scaling analysis. The inserted histogram plot illustrates the distribution of ELM2 simulations.

LH shows high values over the Central and Southeast, and lower values over the West and Southwest (Figure 3b). ELR generally shows higher values over regions with high surface temperature in the South (Figure 3c). ASR shows higher values over the southwestern regions determined by incoming solar radiation and albedo (Figure 3d). Driven by spatial heterogeneity of the 1-km resolution input data, the k-scale simulation captures spatial variability of water and energy fluxes that is not possible by modeling at coarser resolutions and/or using input data at coarser resolutions. Modeling this fine-scale spatial variability is important for modeling the exchange of energy and water fluxes between the land surface and atmosphere.

The relationships between spatial variabilities and spatial scales for SM is demonstrated for two locations L1-2 shown in Figure 3a. We developed a regression relationship $\log\left(\frac{\sigma_{scale}}{\sigma_{1\,km}}\right) \propto \beta \times \log\left(\frac{\lambda_{scale}}{\lambda_{1\,km}}\right)$, where β is an indicator used to quantify the change in spatial variability across scales (Hu et al. 1997). A more negative β indicates a larger dependency of the spatial variability on spatial scales, resulting in a higher information loss (γ). The two locations are specifically chosen to showcase varying levels of spatial information loss with L1 and L2 demonstrating a relatively large and small loss, respectively. At location L1 (Figure 4a), when the 1-km simulation is upscaled to coarser resolutions (i.e., larger spatial scale ratios), the spatial variability of SM decreases, resulting in a negative slope of β.

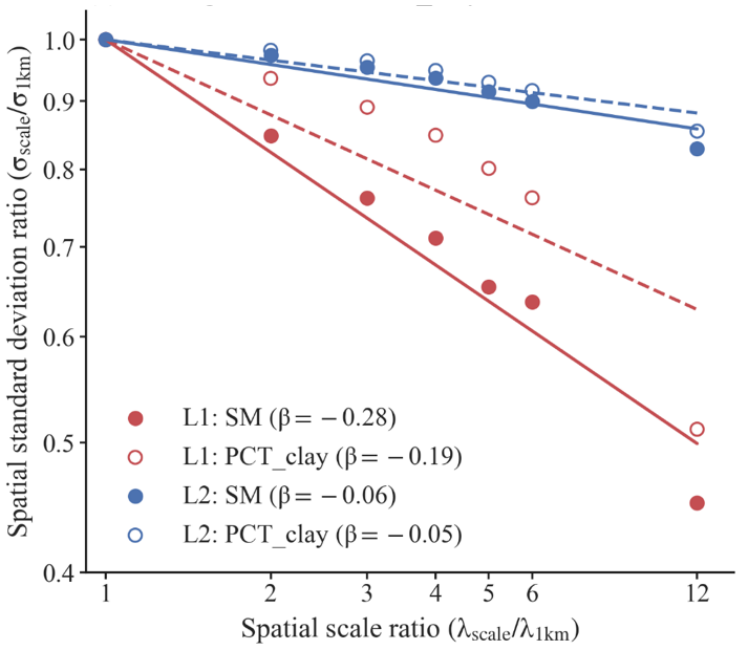

Figure 4. The scaling of spatial variabilities for soil moisture (SM) and percent clay (PCT_CLAY). The slope of the linear regression line, β, quantifies the strength of the negative relationship between spatial scale and spatial variability. A more negative β value indicates a higher spatial-scale dependency and increased information loss at coarser spatial scales. The 0.5° × 0.5° boxes (displayed in Figure 4a), namely L1 and L2, are chosen to contrast larger and smaller negative β values for SM and percent clay.

As shown in Figure 5a, compared to the original 1-km resolution, the information loss γ reaches up to 54.9% at the 12-km spatial scale. The spatial pattern of SM is consistent with the spatial pattern of percent clay (Figure 5a versus 5b and 5c versus 5d), indicating that soil texture contributes significantly to the spatial variability of SM. However, SM has a more negative β than the percent clay (β = –0.28 versus –0.19 at L1, as shown in Figure 5a), suggesting that SM variability is amplified likely by other processes that are also influenced by soil texture. In contrast to location L1, location L2 exhibits less negative β values for both SM and percent clay, suggesting that their spatial variabilities exhibit less scale dependence (Figures 5a, 6c, and 6d). Both SM and percent clay at location L2 approximately maintain their spatial patterns of high values in the west and low values in the east across spatial scales (Figure 5c and 5d).

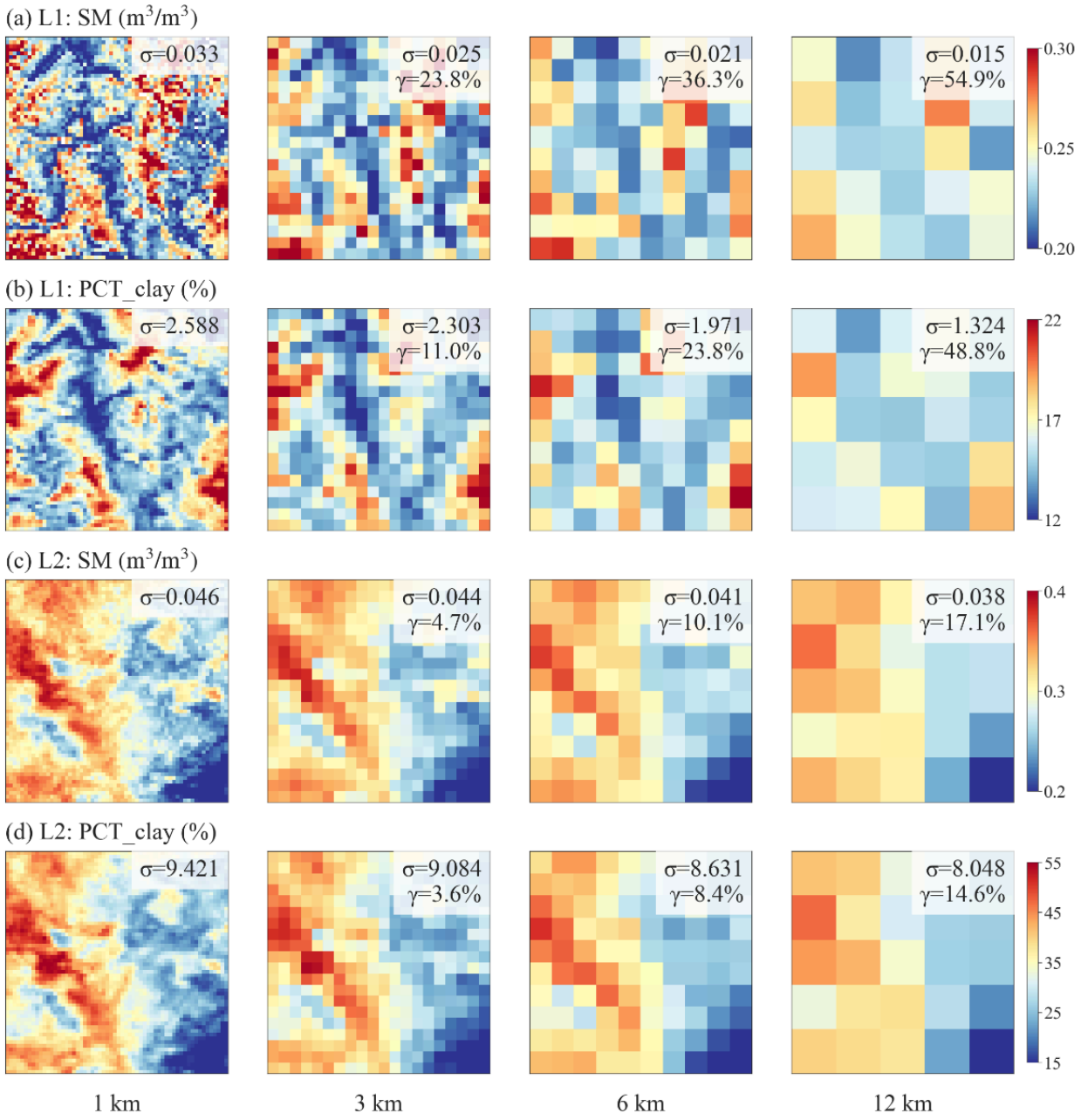

Figure 5. Comparison of soil moisture (SM) and percent clay (PCT_CLAY) across spatial scales at locations L1 and L2 highlighted in Figure 4. Each subplot displays the spatial patterns of SM or percent clay within a 0.5° × 0.5° box, with the spatial standard deviation, σ, and information loss, γ, presented in the legend.

To measure the spatial variability simulated at 1-km resolution, AI/ML techniques were used. The spatial variability was quantified using the standard deviation (σ) within each 0.5° × 0.5° box across CONUS. Four ML models based on eXtreme Gradient Boosting (XGBoost; Chen and Guestrin 2016) were built to explore the spatial relationships between σ and its potential drivers including σ and the spatial mean (μ) of the land surface parameters and the temperature and precipitation averaged over the grid box. Overall, the ML models performed well in predicting the σ of the simulated variables, with small Root Mean Square Error (RMSE) and large correlation coefficient (R^2). SM shows larger spatial

variability in the U.S. Southern Coastal Plain, lower Mississippi River, Northeast, Southeast, and regions around the Great Lakes (Figure 6a), which is roughly consistent with the spatial heterogeneity of the high-resolution SM simulation in Vergopolan et al. (2022). Based on analysis using the game theoretic approach SHapley Additive exPlanations (SHAP; Lundberg and Lee 2017, Lundberg et al. 2018, 2020), the spatial variability of SM across CONUS is driven by various factors, mainly including the spatial variabilities of percent sand and percent clay, mean precipitation, the σ and μ of soil organic matter, the σ of canopy height, and mean temperature (Figure 6b).

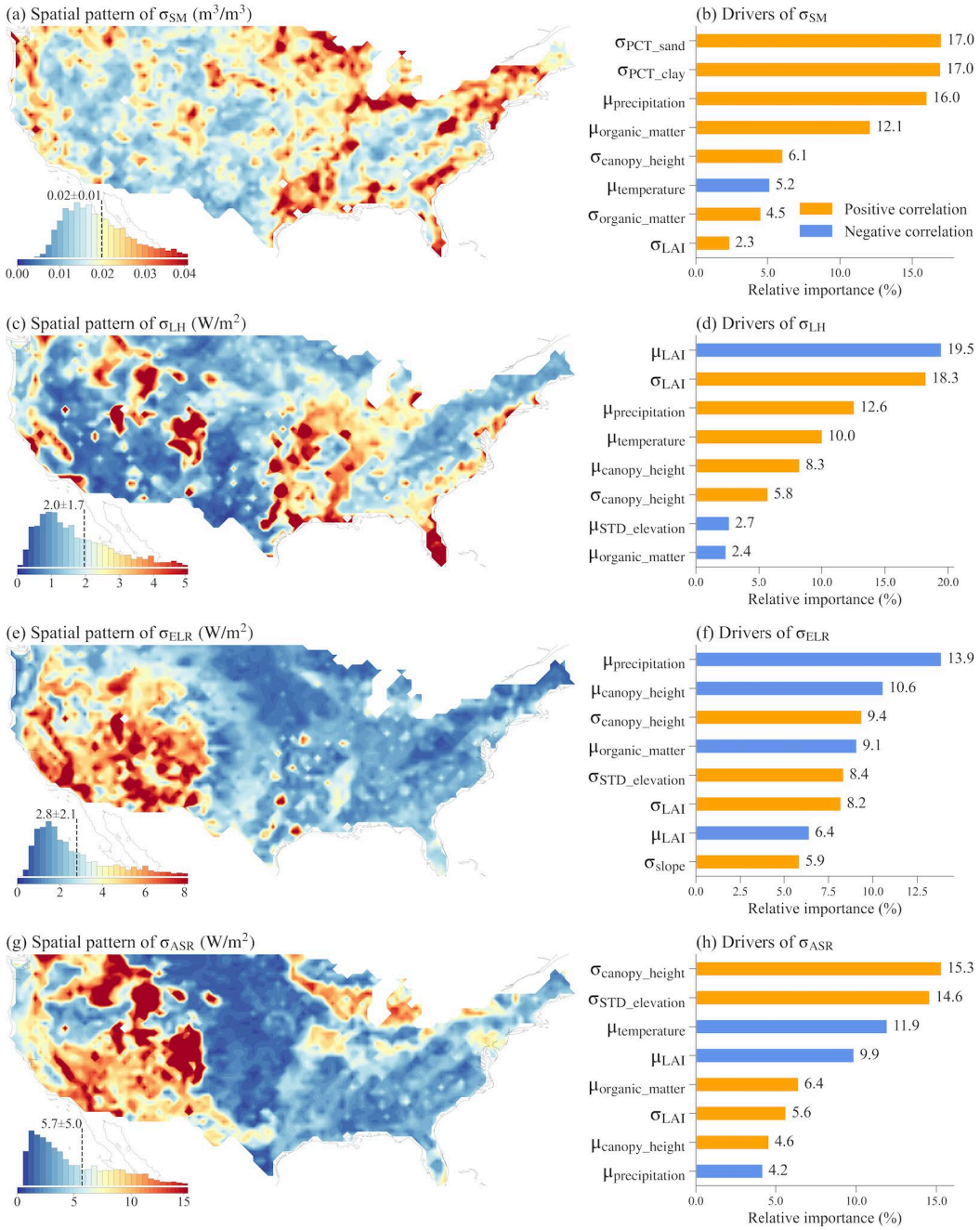

Figure 6. The spatial variability, σ, over each 0.5º × 0.5º grid cell (left plots) and the top eight most important drivers (right plots) of the spatial variability for soil moisture (SM), latent heat flux (LH),

emitted longwave radiation (ELR), and absorbed shortwave radiation (ASR). The inserted histogram plot illustrates the probability distribution of the spatial variability across CONUS. The relative importance of each variable in determining the spatial variability is calculated as the ratio of the mean |SHAP value| of the variable to the sum of the mean |SHAP value| of all variables. Therefore, the sum of the relative importance of all variables is 100%. The yellow and the blue color in the horizontal histograms for the top eight most important drivers indicate a positive and a negative correlation, respectively.

The spatial variability of LH is large in the southeastern, central, and western mountainous regions of the U.S. (Figure 6c). Vegetation properties and climate conditions mainly drive the variability of LH (Figure 6d). The μ and σ of LAI can affect transpiration and soil evaporation, while canopy height can influence surface roughness length and, in turn, evapotranspiration. Mean precipitation and temperature reflect the overall climate conditions related to the water and energy available for latent heat.

ELR and ASR exhibit large spatial variability mainly over the western U.S., with ASR additionally showing significant spatial variability across the northern U.S. (Figure 6e and 6g). This variability is primarily driven by climate conditions such as mean precipitation and temperature, topographic features such as standard deviation of elevation and slope, and vegetation properties including LAI and canopy height (Figure 6f and 6h). These factors are related to the radiation input and surface properties, such as albedo and roughness length, which impact the energy cycles and availability of ELR and ASR.

3.3 Improvement in the Prediction of Water Table Depth in the 1-km ELM CONUS Simulation

As ELM simulations were typically performed at resolutions of 25-100 km, applying ELM at k scale may not yield skillful simulations without recalibrating the ELM parameterizations for k-scale applications. The ELM-simulated results presented in Section 3.2 used the default, spatially homogeneous values of the subsurface runoff generation parameter (f_{drain}) calibrated for coarse-scale simulations. However, the estimated optimal f_{drain} to improve the simulation of water dynamics shows significant spatial variability (Figure 7). For 51% of all the grid cells within the 1 km CONUS domain, the optimal f_{drain} value resulted in WTD that matched the F2013 data set. For 31% of all the grid cells, ELMv2 could not simulate WTD that was shallow enough to match the F2013 data set, while for the remaining 18% of all the grid cells, the ELMv2-simlated WTD was not deep enough to match the F2013 data set.

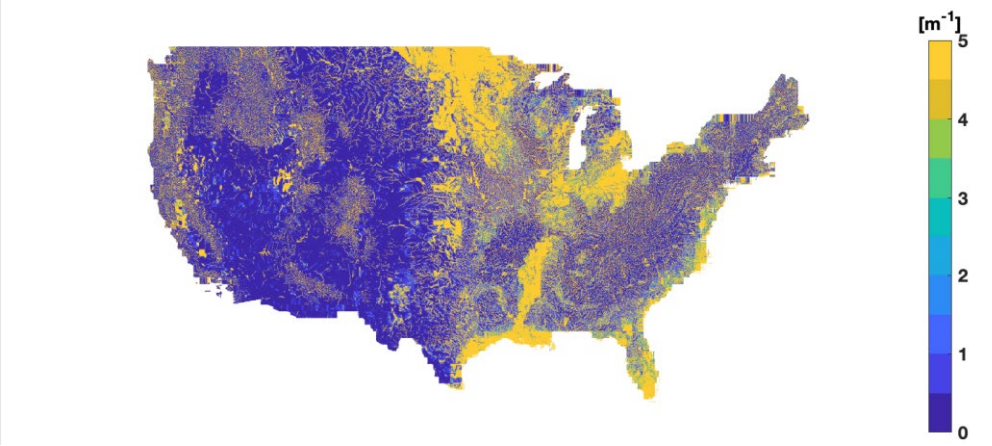

Figure 7. The calibrated values of ELMv2's subsurface runoff parameter, f_{drain}.

The ELM-simulated WTD using the default f_{drain} value (Figure 8a) shows low spatial variability compared to the F2013 data set (Figure 8c) and is unable to predict WTD deeper than 15 m. On the other hand, the ELMv2 simulation of WTD using the calibrated f_{drain} shows significant agreement with the F2013 data set (Figure 8b). Limiting analysis to grid cells with WTD shallower than 40 m in the F2013 data set since the depth of the ELM soil column is ~40 m, the model shows improvement in all three statistical metrics with the use of calibrated f_{drain}. The bias, RMSE, and R2 improve from 5.2 m to -1.7 m, from 12.4 m to 3.2 m, and from 0.3 to 0.98, respectively. Although the calibrated f_{drain} can significantly improve the simulation of WTD, other ELM parameters need to be calibrated to further improve the prediction of the shallower WTD.

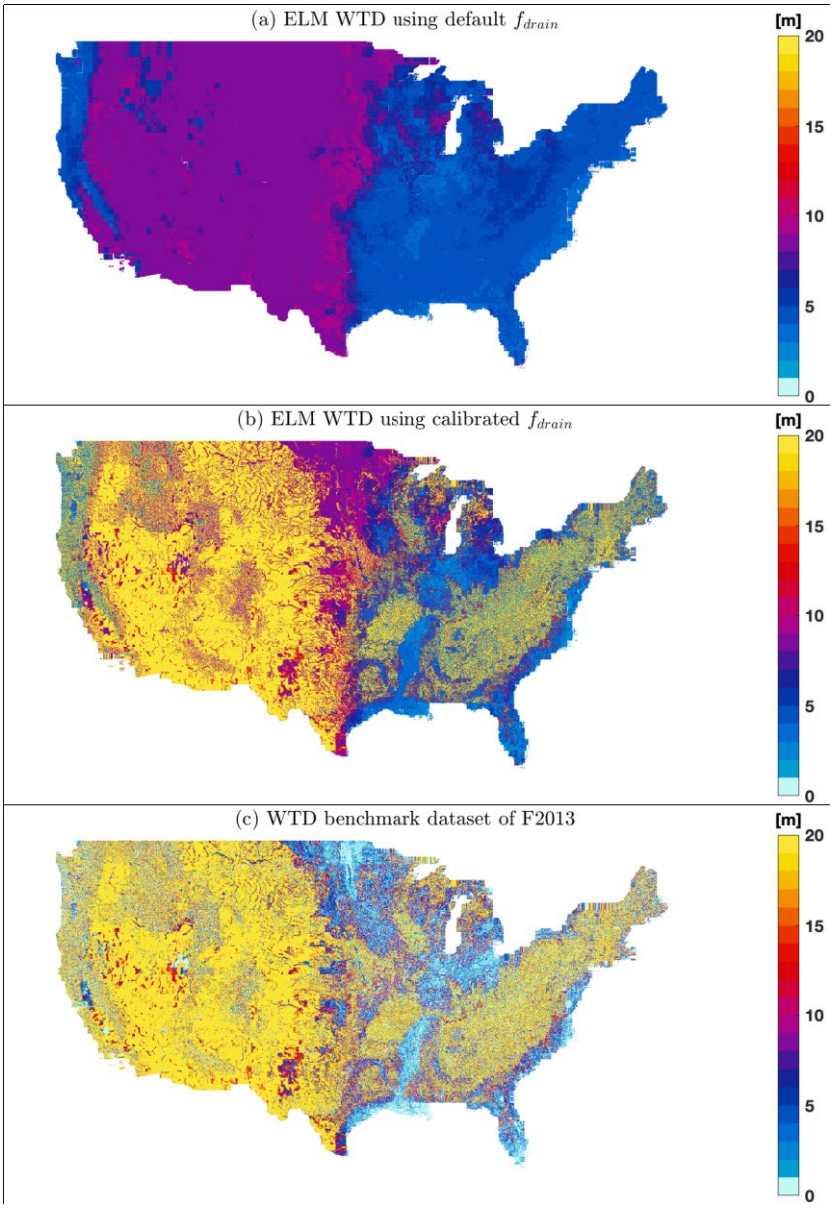

Figure 8. Water table depth (WTD) simulated by ELM using (a) the spatially homogenous, default subsurface runoff parameter, f_{drain}, and (b) the calibrated f_{drain} (shown in Figure 7), and (c) the WTD benchmark data set of Fan et al. (2013).

4.0 Summary and Future Work

This report documents three major accomplishments for high-resolution land surface modeling. First, global 1-km land surface parameter data sets have been developed that represent significant improvements over the current data sets. Compared to the two common land surface parameter data sets (Lawrence et al. 2019, Ke et al. 2012), the new data are more advanced by using the newest high-resolution data sources, including MODIS PFTs and non-vegetation land units, LAI and SAI, canopy height, soil properties, and topography factors. The availability of multi-year data for LULC, LAI, and SAI parameters is advantageous for studies such as LULC changes, including urbanization, deforestation, and agricultural impacts. Second, land surface parameters are shown to contribute to significant spatial heterogeneity in ELM2 simulations of soil moisture, latent heat, emitted longwave radiation, and absorbed shortwave radiation. On average, about 31% to 54% of spatial information is lost by upscaling the 1-km ELM2 simulations to a 12-km resolution. The XML analysis reveals that the spatial variability and spatial information loss of ELM2 simulations are primarily impacted by the spatial variability and information loss of soil properties, vegetation properties, and topography factors, as well as the mean climate conditions. Third, significant improvements in the simulation of water table over CONUS at 1 km is achieved through model calibration.

The new land simulation capability presents multiple opportunities for k-scale Earth system modeling. First, the new input data sets can be used to improve the representation of land-atmosphere interactions in the Simple Cloud-Resolving E3SM Atmosphere Model (SCREAM) (Caldwell et al. 2021), as well as the Multiple Atmosphere Multiple Land approach used in multiscale modeling framework such as MMF-E3SM (Baker et al. 2019, Lin et al. 2023, Lee et al. 2023). Second, k-scale offline LSMs can be used to study terrestrial processes (e.g., surface energy fluxes, soil moisture dynamics, and biogeochemical cycle) and their interactions, and be evaluated against k-scale benchmark data sets. Lastly, the new k-scale land simulation capability will further facilitate modeling of human-Earth interactions to support the use of E3SM in addressing energy-related questions.

5.0 References

Baker, IT, AS Denning, DA Dazlich, AB Harper, MD Branson, DA Randall, MC Phillips, KD Haynes, and SM Gallup. 2019. “Surface-Atmosphere Coupling Scale, the Fate of Water, and Ecophysiological Function in a Brazilian Forest.” *Journal of Advances in Modeling Earth Systems* 11(8): 2523–2546, https://doi.org/10.1029/2019ms001650

Bisht, G, WJ Riley, HM Wainwright, B Dafflon, F Yuan, and VE Romanovsky. 2018. “Impacts of microtopographic snow redistribution and lateral subsurface processes on hydrologic and thermal states in an Arctic polygonal ground ecosystem: a case study using ELM-3D v1. 0.” *Geoscientific Model Development* 11(1): 61–76, https://doi.org/10.5194/gmd-11-61-2018

Caldwell, PM, CR Terai, B Hillman, ND Keen, P Bogenschutz, W Lin, H Beydoun, M Taylor, L Bertagna, AM Bradley, and TC Clevenger. 2021. “Convection-permitting simulations with the E3SM global atmosphere model.” *Journal of Advances in Modeling Earth Systems* 13(11): e2021MS002544, https://doi.org/10.1029/2021MS002544

Chen, T, and C Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 785–794, https://doi.org/10.1145/2939672.2939785

Condon, LE, AL Atchley, and RM Maxwell. 2020. "Evapotranspiration depletes groundwater under warming over the contiguous United States." *Nature Communications* 11(1): 873, https://doi.org/10.1038/s41467-020-14688-0

Döll, P. 2009. "Vulnerability to the impact of climate change on renewable groundwater resources: a global-scale assessment." *Environmental Research Letters* 4(3): 035006, https://doi.org/10.1088/1748-9326/4/3/035006

Fan, Y, H Li, and G Miguez-Macho. 2013. "Global patterns of groundwater table depth." *Science* 339(6122): 940–943, https://doi.org/10.1126/science.1229881

Friedl, M, and D Sulla-Menashe. 2019. MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed 2022-11-21 from https://doi.org/10.5067/MODIS/MCD12Q1.006

Hu, Z, S Islam, and Y Cheng. 1997. "Statistical characterization of remotely sensed soil moisture images." *Remote Sensing of Environment* 61(2): 310–318, https://doi.org/10.1016/s0034-4257(97)89498-9

Ke, Y, LR Leung, M Huang, AM Coleman, H Li, and MS Wigmosta. 2012. "Development of high resolution land surface parameters for the Community Land Model." *Geoscientific Model Development* 5(6): 1341–1362, https://doi.org/10.5194/gmd-5-1341-2012

Kim, H. 2017. Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1) [Data set]. Data Integration and Analysis System (DIAS). https://doi.org/10.20783/DIAS.501

Lang, N, W Jetz, K Schindler, and JD Wegner. 2023. "A high-resolution canopy height model of the Earth." *Nature Ecology and Evolution* 7: 1778–1789, https://doi.org/10.1038/s41559-023-02206-6

Lawrence, DM, RA Fisher, CD Koven, KW Oleson, SC Swenson, G Bonan, N Collier, B Ghimire, L van Kampenhout, D Kennedy, E Kluzek, PJ Lawrence, F Li, H Li, D Lombardozzi, WJ Riley, WJSacks, M Shi, M Vertenstein, WR Wieder, C Xu, AA Ali, AM Badger, G Bisht, M van den Broeke, MA Brunke, SP Burns, J Buzan, M Clark, A Craig, K Dahlin, B Drewniak, JB Fisher, M Flanner, AM Fox P Gentine, F Hoffman, G Keppel-Aleks, R Knox, S Kumar, J Lenaerts, LR Leung, WH Lipscomb, Y Lu, A Pandey, JD Pelletier, J Perket, JT Randerson, DM Ricciuto, BM Sanderson, A Slater, ZM Subin, J Tang, RQ Thomas, M Val Martin, and X Zeng. 2019. "The Community Land Model Version 5: Description of New Features, Benchmarking, and Impact of Forcing Uncertainty." *Journal of Advances in Modeling Earth Systems* 11(12): 4245–4287, https://doi.org/10.1029/2018ms001583

Lee, J, WM Hannah, and DC Bader. 2023. "Representation of atmosphere induced heterogeneity in land-atmosphere interactions in E3SM-MMFv2." *Geoscientific Model Development* 16(24): 7275–7287, https://doi.org/10.5194/gmd-16-7275-2023

Leung, LR, DC Bader, MA Taylor, and RB McCoy. 2020. "An Introduction to the E3SM Special Collection: Goals, Science Drivers, Development, and Analysis." *Journal of Advances in Modeling Earth Systems* 12(11): e2019MS001821, https://doi.org/10.1029/2019ms001821

Li, L, G Bisht, D Hao, and L-YR Leung. 2023. "Global 1km Land Surface Parameters for Kilometer-Scale Earth System Modeling." *Earth System Science Data*. Preprint: in review. https://doi.org/10.5194/essd-2023-242

Lin, G, LR Leung, J Lee, BE Harrop, IT Baker, MD Branson, AS Denning, CR Jones, M Ovchinnikov, DA Randall, and Z Yang. 2023. "Modeling Land-Atmosphere Coupling at Cloud-Resolving Scale within the Multiple Atmosphere Multiple Land (MAML) Framework in SP-E3SM." *Journal of Advances in Modeling Earth Systems* 15(2): e2022MS003101, https://doi.org/10.1029/2022ms003101

Lundberg, S, and S-I Lee. 2017. "A Unified Approach to Interpreting Model Predictions." *arXiv* 1705.07874[cs.AI], https://doi.org/10.48550/arXiv.1705.07874

Lundberg, SM, B Nair, MS Vavilala, M Horibe, MJ Eisses, T Adams, DE Liston, DK-W Low, S-F Newman, J Kim, and S-I Lee. 2018. "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery." *Nature Biomedical Engineering* 2: 749–760, https://doi.org/10.1038/s41551-018-0304-0

Lundberg, SM, G Erion, H Chen, A DeGrave, JM Prutkin, B Nair, R Katz, J Himmelfarb, N Bansal, and S-I Lee. 2020. "From local explanations to global understanding with explainable AI for trees." *Nature Machine Intelligence* 2: 56–67, https://doi.org/10.1038/s42256-019-0138-9

Maxwell, RM, and SJ Kollet. 2008. "Interdependence of groundwater dynamics and land-energy feedbacks under climate change." *Nature Geoscience* 1(10): 665–669, https://doi.org/10.1038/ngeo315

Poggio, L, LM de Sousa, NH Batjes, GBM Heuvelink, B Kempen, E Ribeiro, and D Rossiter. 2021. "SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty." *Soil* 7(1): 217–240, https://doi.org/10.5194/soil-7-217-2021

Ruiz-Vásquez, M, S O, G Arduini, S Boussetta, A Brenning, A Bastos, S Koirala, G Balsamo, M Reichstein, and R Orth. 2023. "Impact of updating vegetation information on land surface model performance." *Journal of Geophysical Research – Atmospheres* 128(21): e2023JD039076, https://doi.org/10.1029/2023JD039076

Slingo, J, P Bates, P Bauer, S Belcher, T Palmer, G Stephens, B Stevens, T Stocker, and G Teutsch. 2022. "Ambitious partnership needed for reliable climate prediction." *Nature Climate Change* 12: 499–503, https://doi.org/10.1038/s41558-022-01384-8

Vergopolan, N, J Sheffield, NW Chaney, M Pan, HE Beck, CR Ferguson, L Torres-Rojas, F Eigenbrod, W Crow, and EF Wood. 2022. "High-Resolution Soil Moisture Data Reveal Complex Multi-Scale Spatial Variability across the United States." *Geophysical Research Letters* 49(15): e2022GL098586, https://doi.org/10.1029/2022gl098586

Xu, D, G Bisht, Z Tan, C Liao, T Zhou, H-Y Li, and L-YR Leung. 2023. "Disentangling the Hydrological and Hydraulic Controls on Streamflow Variability in E3SM V2 – A Case Study in the Pantanal Region." *EGUsphere* (preprint), https://doi.org/10.5194/egusphere-2023-1879

Yamazaki, D, D Ikeshima, J Sosa, PD Bates, GH Allen, and TM Pavelsky. 2019. "MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset." *Water Resources Research* 55(6): 5053–5073, https://doi.org/10.1029/2019wr024873

Yuan, H, Y Dai, Z Xiao, D Ji, and W Shangguan. 2011. "Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling." *Remote Sensing of Environment* 115(5): 1171–1187, https://doi.org/10.1016/j.rse.2011.01.001

U.S. DEPARTMENT OF
ENERGY

Office of Science