https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


Open Access Earth System Science Data Discussions

# Global 1km Land Surface Parameters for Kilometer-Scale Earth System Modeling

Lingcheng Li, Gautam Bisht, Dalei Hao, L. Ruby Leung

Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA

Correspondence: Lingcheng Li (lingcheng.li@pnnl.gov)

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


Open Access Earth System Science Data Discussions

**Abstract**

Earth system models (ESMs) are progressively advancing towards the kilometer scale (k-scale). However, the surface parameters for Land Surface Models (LSMs) within ESMs running at the k-scale are typically derived from coarse resolution and outdated datasets. This study aims to develop a new set of global land surface parameters with a resolution of 1 km for multiple years from 2001 to 2020, utilizing the latest and most accurate available datasets. Specifically, the datasets consist of parameters related to land use and land cover, vegetation, soil, and topography. To demonstrate the capability of these new parameters, we conducted 1 km resolution simulations using the E3SM Land Model version 2 (ELM2) over the contiguous United States. Our results demonstrate that land surface parameters 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. Using eXplainable Machine Learning (XML) methods, the influential factors driving the spatial variability and spatial information loss of ELM2 simulations were identified, highlighting the substantial impact of the spatial variability and information loss of various land surface parameters, as well as the mean climate conditions. The new land surface parameters are tailored to meet the emerging needs of k-scale LSMs and ESMs modeling with significant implications for advancing our understanding of water, carbon, and energy cycles under global change. The 1 km land surface parameters are publicly available at https://doi.org/10.25584/PNNLDH/1986308 (Li et al., 2023).

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




## 1. Introduction

Aided by advancements in computing power, it has become increasingly feasible to run land surface models (LSMs) and Earth system models (ESMs) at the kilometer scale (k-scale) to improve our understanding of Earth system processes. The emergence of k-scale modeling has the potential to improve the accuracy of climate simulations significantly and allow for explicit modeling of physical processes that were previously poorly represented in climate models (Nat. Clim. Chang. 2022), such as modeling of mesoscale convective systems in the atmosphere (Slingo et al., 2022) and mesoscale eddies in ocean (Hewitt et al., 2022). Simultaneously, land modeling has also witnessed a surge of interest in hyper-resolution modeling, initially proposed by Wood et al. (2011), which aims to model land surface processes at a horizontal resolution of 1 km globally and 100 m or finer for continental or regional domains. The motivation behind hyper-resolution modeling is to address the requirements of operational forecasting like extreme events, and to enhance our understanding of hydrological and biogeochemical cycling, and land–atmosphere interactions. High-resolution LSMs have been increasingly applied in various fields, as demonstrated by recent examples, such as 30-meter soil moisture simulations over the contiguous United States (CONUS) (Vergopolan et al., 2020, 2021, 2022), 500-meter hyper-resolution modeling of surface and root zone soil moisture over Oklahoma (Rouf et al., 2021), 1-km simulations over Southwestern US (Singh et al., 2015), 3-km simulations over eastern Tibetan Plateau to understand hydrological changes over mountainous regions (Yuan et al., 2018; Ji and Yuan, 2018), 6-km simulations over China to reduce simulations errors of hydrological variables (Ji et al., 2023). High-resolution modeling can better capture the land surface heterogeneity and could improve simulations of terrestrial water, energy, and biogeochemical cycles, as well as land and atmosphere coupling (Giorgi and Avissar, 1997; Chaney et al., 2018; Zhou et al., 2019; Liu et

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




al., 2017; Bou-Zeid et al., 2020; Chen et al., 2020; Nitta et al., 2020; Vrese et al., 2016). Singh et al. (2015) demonstrated that increasingly capturing topography and soil texture heterogeneity at finer resolutions (e.g., 1 km) improves land surface modeling of water and energy variables. Li et al. (2022) have shown that the spatial heterogeneities of land surface parameters (including land use and land cover (LULC) and topography) are essential for modeling the spatial variability of land surface energy and water partitioning. Hao et al. (2022) found that 1 km simulations with sub-grid topographic configurations can better capture the topographic effects on surface fluxes.

The parameters for LSMs within ESMs being run at the k-scale are typically derived from coarse resolution datasets or outdated datasets. 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 utilized. Publicly available land surface parameters are primarily provided at coarse resolutions and based on outdated datasets. For example, the Community Land Model version 5 (CLM5; Lawrence et al., 2019) typically relies on land surface parameters with spatial resolutions ranging from 1km to 0.5º based on source datasets that were processed more than 10 years ago (see Table 1 for details). Although LULC-related parameters are available at a relatively high resolution of 0.05º, they are temporally static and were derived from a combination of data from different years spanning 1993 to 2012 (Table 1). Leaf area index (LAI) was derived from the now outdated products of Moderate Resolution Imaging Spectroradiometer (MODIS) collection 4 (Myneni et al., 2002). The canopy height for tree Plant Functional Types (PFTs) is based on forest canopy height data derived from the Geoscience Laser Altimeter System (GLAS) aboard ICESat, collected in 2005 (Simard et al., 2011). Canopy height for short vegetation is represented by PFT-specific values that remain invariant in space (Bonan et al., 2002). Soil sand

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




and clay content were obtained from the International Geosphere-Biosphere Programme (IGBP) soil dataset (Global Soil Data Task 2000) consisting of 4931 soil mapping units (IGBP, 2000). These CLM5 land surface parameters have been widely utilized in the LSMs and ESMs community, despite being developed over a decade ago. Subsequently, Ke et al. (2012; hereafter referred to as K2012) developed an updated set of LULC and vegetation-related land surface parameters for CLM4 at a resolution of 0.05º. These parameters were developed based on MODIS collection 5 products or datasets derived from MODIS collection 5 products, including PFTs and non-vegetation land cover, LAI, and Stem Area Index (SAI). K2012 has also been widely used by LSMs, including CLM (e.g., Leng et al., 2013; Singh et al., 2015; Ke et al., 2013; Xia et al., 2017) and the Energy Exascale Earth System Model (E3SM) Land Model (ELM) (e.g., Caldwell et al., 2019; Leung et al., 2020; Li et al., 2022). However, the CLM5 and K2012 datasets, with their relatively coarse resolution and reliance on outdated data from over a decade ago, may not fully meet the requirements for k-scale modeling. Additionally, these datasets include LULC, LAI, and SAI that are year invariant. Consequently, they are inappropriate for studies involving LULC changes, such as urbanization. In addition, some recently developed land surface processes and their associated parameters are not included in previous datasets. For instance, Hao et al. (2021) introduced a sub-grid topographic parameterization of solar radiation with five associated topographic factors in ELM, which have been found to significantly affect the surface energy budget.

High-resolution and up-to-date datasets at kilometer or finer resolutions are now widely available and can be utilized to derive more accurate land surface parameters for k-scale LSM simulations. For example, the MODIS Land Cover Type Collection 6 (MCD12Q1 C6) data product provides

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




97 global land cover types yearly from 2001 to the present (Friedl et al., 2019; Sulla-Menashe et al.,
98 2019) at 500-meter resolution. Compared to the MODIS Collection 4 (used in CLM5 land surface
99 parameters) and Collection 5 products (used in K2012 land surface parameters), the C6 data
100 represents a significant advancement in algorithm improvements and the quality of land cover
101 information. Despite the availability of high-resolution MODIS LAI products, such as the 500 m
102 MCD15A2H (Myneni et al., 2021), they suffer from noise and gaps with spatially and temporally
103 inconsistent values due to clouds, seasonal snow cover, instrument issues, and uncertainties in
104 retrieval algorithms (Yuan et al., 2011). To address these limitations, Yuan et al. (2011)
105 reprocessed MODIS LAI products and generated a more accurate and spatiotemporally continuous
106 and consistent LAI dataset that is available continuously to the present period. Additional high-
107 resolution and up-to-date datasets are available for preparing land surface parameters, such as soil
108 texture and soil organic matter at 250-meter resolution (Poggio et al., 2021) and vegetation height
109 at 10-m resolution (Lang et al., 2022).
110
111 This study aims to develop a new set of global land surface parameters with a resolution of 1 km
112 for multiple years, utilizing the latest and most accurate available datasets. These parameters will
113 be tailored to meet the needs of k-scale Earth system modeling. The newly developed land surface
114 parameters include four categories: (1) LULC-related parameters, such as the spatial distributions
115 of PFTs, lakes, wetlands, urban areas, and glaciers; (2) vegetation-related parameters, including
116 PFTs' LAI and SAI for multiple years ranging from 2001 to 2021, and the canopy top and bottom
117 height; (3) soil-related parameters, such as soil textures and soil organic matter; and (4)
118 topography-related parameters, such as slope, aspect, and sub-grid topographic factors. We
119 employed the ELM version 2 (ELM2) as a testbed to demonstrate the capability of the new high-

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


Open Access Earth System Science Data Discussions

120 resolution parameters by conducting a 5-year 1 km resolution simulation over the CONUS. We
121 performed a spatial scaling analysis on four ELM2 simulated variables, which included soil
122 moisture, latent heat, emitted longwave radiation, and absorbed shortwave radiation, to underscore
123 the significance of high-resolution land surface parameters on ELM2 simulations. We employed
124 eXplainable Machine Learning (XML) methods to evaluate the most important factors of land
125 surface parameters and climate conditions (e.g., mean temperature and precipitation) in driving the
126 spatial variability and spatial information loss of ELM2 simulations.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


Open Access Earth System Science Data Discussions

## 2. Development of 1km land surface parameters

In this study, all the land surface parameters were developed globally at a resolution of approximately 1 km (i.e., 1/120°, hereafter referred to as 1 km; Table 1). The LULC-related parameters, soil properties, canopy height, and elevation were processed via Google Earth Engine (GEE; Gorelick et al., 2017). The LAI was processed using an area-weighted average from its original 450 m resolution obtained from Beijing Normal University (Yuan et al., 2011). The detailed methods for deriving these parameters are described below.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




135 Table 1 Comparison between new and previous land surface parameters

| Category | Land surface parameters | This study | ELM2 / CLM5 * | K2012 |
|---|---|---|---|---|
| LULC | PFTs,<br>Lake,<br>Glacier,<br>Urban | • Resolution: 1 km, yearly, 2001-2020<br>• Data source: 500 m, yearly, MODIS collection 6 (Friedl et al., 2022) | • Resolution: 0.05°, temporally static, processed based on data from mixed years<br>• PFTs data source: mixed years from 1993 to 2001; 500 m, MODIS Vegetation Continuous Fields (Hansen et al., 2003); 1 km, tree cover (Defries et al., 2000); 10 km (5 arc minutes), cropland (Ramankutty and Foley, 1999); 1 km, MODIS land cover collection 4 (Friedl et al., 2002)<br>• Lake data source: 3 km (90 arc seconds) lake data (Kourzeneva 2009, 2010)<br>• Glacier data source: glacier and ice sheet vector data (Arendt et al. 2012; Rastner et al. 2012)<br>• Urban data source: 1 km urban data (Jackson et al., 2010) | • Resolution: 0.05°, year 2005<br>• Data source: 500 m, yearly, MODIS collection 5 (Friedl et al., 2010) |
| Vegetation | LAI,<br>SAI | • Resolution: 1 km, monthly, 2001-2020<br>• Data source: 450 m, 8-day, reprocessed MODIS collection 6 LAI (Yuan et al., 2011; Friedl et al., 2022) | • Resolution: 0.5°, 12 months<br>• Data source: 1 km, 8-day, MODIS collection 4 LAI (Myneni et al., 2002) | • Resolution: 0.05°, year 2005<br>• Data source: 450 m, 8-day, reprocessed MODIS collection 5 LAI (Yuan et al., 2011; Friedl et al., 2010) |
| | Canopy top height,<br>Canopy bottom height | • Resolution: 1 km, temporally static<br>• Data source: 10 m, vegetation canopy height (Lang et al., 2022) | • Resolution: 0.5° or PFT specified value, temporally static<br>• Tree PFT data source: 1 km, forest canopy height derived using 2005 GLAS aboard ICESat data (Simard et al., 2011);<br>• Short vegetation data source: PFT specific values (Bonan et al., 2002) | -- |
| Soil | Percent sand,<br>Percent clay | • Resolution: 1 km, temporally static<br>• Data source: 250 m, Soilgrid v2 (Poggio et al., 2021) | • Resolution: 10 km (0.083°), temporally static<br>• Data source: IGBP soil data of 4931 mapping units (IGBP, 2000) | -- |
| | Soil organic matter | | | |
| Topography | Elevation | • Resolution: 1 km, temporally static<br>• Data source: 90 m, MERIT Hydro elevation (Yamazaki et al., 2019) | • Resolution: 1 km, temporally static<br>• Data source: USGS HYDRO1k (Verdin and Greenlee 1996) | -- |
| | Slope | | | |
| | Standard deviation of elevation | • Resolution: 1 km, temporally static<br>• Data source: 90 m, Hydro elevation (Yamazaki et al., 2019) | -- | |
| | Aspect | | | |
| | Sky view factor | | | |
| | Terrain view factor | | | |

136 * ELM2 and CLM5 share the same default land surface parameters, detailed descriptions available at:
137 https://escomp.github.io/ctsm-docs/versions/release-clm5.0/html/tech_note/index.html.

138
139

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




## 2.1 LULC-related parameters

This study utilized MODIS MCD12Q1 version 6 (Friedl et al., 2022) to derive the PFT and other non-vegetation land types at a resolution of 1 km for 2001–2020. The original MODIS land cover data was first resampled to 1 km from its original 500 m resolution using a majority resampling method in GEE. At such a high 1km resolution, we did not consider the proportion of different land cover types within each grid. Instead, we assigned 100% of a grid cell to the major land cover type. Specifically, the MCD12Q1 LC_Type 5 PFT classification layer was used to determine the distributions of the seven PFTs, as well as lake, urban, and glacier, following the method outlined in Ke et al. (2012) and summarized below:

- The seven PFTs include needleleaf evergreen trees (NET), needleleaf deciduous trees (NDT), broadleaf evergreen trees (BET), broadleaf deciduous trees (BDT), shrub (SHR), grass (GRS), and crop (CRO). These PFTs were further reclassified into 15 categories (Table S1) that are typically used in LSMs based on the rules presented in Bonan et al. (2002a) with the assistance of 1 km precipitation and surface air temperature from WorldClim V1 (Hijmans et al., 2005).
- Grass was reclassified as C3 and C4 grass using the approach presented by Still et al. (2003), with the assistance of monthly LAI (processed in section 2.2.1) and meteorological variables from WorldClim V1.
- The "non-vegetated land" was classified as barren soil class.
- The "permanent snow and ice" was assigned as the glacier land unit. Global lakes were identified based on the classification of "water bodies" over the global land, constrained using the global land mask obtained from Natural Earth (https://www.naturalearthdata.com/).

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


- The urban land unit was determined based on the MODIS "urban and built-up" classification. These urban grids were further classified into three urban classes, namely, tall building district (TBD), high density (HD), and medium density (MD), based on Jackson et al. (2010; hereinafter referred to as J2010). J2010 generated global urban extent maps for the TBD, HD, and MD classes at a spatial resolution of 1 km, based on rules of building height and vegetation coverage fraction (https://gdex.ucar.edu/dataset/188a_oleson/file.html). However, the J2010 dataset is temporally static and cannot reflect changes in urban boundaries over time. Therefore, we reclassified the yearly MODIS urban land class as TBD, HD, and MD based on the J2010 dataset using the nearest neighbor sampling method for each year.

After determining the distribution of 15 PFTs, bare soil, lake, glacier, and urban land, any remaining 1 km grids were assigned as ocean (Table S1). It should be noted that the wetland land unit was not explicitly classified in this study. This is because, instead of treating wetlands as an individual land unit, many LSMs (e.g., ELM2 and CLM5) integrate wetland functioning processes prognostically within other land units where a surface water storage component is implemented to represent wetland functioning.

## 2.2 Vegetation-related parameters

### 2.2.1 Monthly LAI and SAI

The monthly LAI parameters were obtained from Beijing Normal University (BNU_LAI; Yuan et al., 2011). BNU_LAI is a reprocessed version of the MODIS LAI C6 product, which has undergone comprehensive quality control and use of multiple algorithms (Yuan et al., 2011). The data have better performance in validation against reference LAI and are more spatiotemporally

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


186 continuous and consistent than the original MODIS LAI (Yuan et al., 2011). The 8-day BNU_LAI
187 product at a resolution of 15 seconds (~450 m) over 2001–2020 was downloaded from
188 http://globalchange.bnu.edu.cn/research/laiv061. Subsequently, the data were resampled to a
189 resolution of 1 km using an area-weighted average method and averaged temporally for each
190 month. The processed monthly LAI at 1 km resolution was subsequently assigned to each of the
191 15 PFTs described above at each grid. The monthly SAI was then calculated based on the
192 processed monthly LAI using the methods and PFT parameters described in Zeng et al. (2002).
193
194 **2.2.2 Vegetation canopy height**
195 The global vegetation canopy height dataset used in this study was obtained from Lang et al. (2022).
196 Lang et al. (2022) developed a probabilistic deep learning model to retrieve canopy height from
197 the Sentinel-2 images by fusing Global Ecosystem Dynamics Investigation (GEDI). This dataset
198 is the first globally consistent and wall-to-wall canopy height at a 10 m spatial resolution and
199 includes canopy height for all vegetation types. The canopy height served as the canopy top height
200 parameter. Canopy bottom height was calculated by multiplying PFT-based ratios derived from
201 the ratio of ELM's (same as CLM5) canopy top and bottom heights for different PFTs (Table S2).
202
203 **2.3 Soil-related parameters**
204 We obtained the Soilgrid v2 data with an original resolution of 250 m (Poggio et al., 2021) to
205 prepare soil properties. Soilgrid is generated using machine learning based on multiple data
206 sources of soil profiles and remote sensing data (Hengl et al., 2017). Soilgrid v2 provides percent
207 clay, percent sand, and soil organic matter for six soil layers: 0–5 cm, 5–15 cm, 15–30 cm, 30–60
208 cm, 60–100 cm, and 100–200 cm. The original SoilGrid version 2 data obtained from GEE were

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


processed at 1 km resolution with multiple layers using an area-weighted average method. To facilitate the demonstration, we restructured the six soil layers vertically into ELM's ten effective soil layers (0–1.8 cm, 1.8–4.5 cm, 4.5–9.1 cm, 9.1–16.6 cm, 16.6–28.9 cm, 28.9–49.3 cm, 49.3–82.9 cm, 82.9–138.3 cm, 138.3–229.6 cm, and 229.6–380.2 cm) using the nearest neighboring method. It should be noted that the lake module in ELM2 and CLM5 requires soil properties, but the Soilgrid v2 data may not provide coverage over water surfaces. To address this, we utilized the nearest neighbor sampling method to map the 1 km soil properties onto the terrestrial water surface.

**2.4 Topography-related parameters**

The 90 m digital elevation from MERIT Hydro (Yamazaki et al., 2019) was used to derive topography-related parameters. We first acquired the 1km elevation and standard deviation of elevation using GEE based on the original 90 m elevation. Further, we calculated the slope, aspect, sky view factor, and terrain configuration factor from the 1km elevation using the parallel computing tool developed by Dozier (2022). The sky view factor represents the proportion of visible sky limited by adjacent terrain, and the terrain configuration factor describes the proportion of adjacent terrain which is visible to the ground target. Finally, to drive the parameterization of sub-grid topographical effects on solar radiation (Hao et al., 2022) in ELM2, we calculated the $\sin(slope) \cdot \sin(aspect)$ and $\sin(slope) \cdot \cos(aspect)$ for calculating the local solar incident angle, and two normalized angle-related factors, the sky view factor, and terrain configuration factor by $\cos(slope)$.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




## 3. K-scale demonstration simulation over CONUS

### 3.1 Experiment design

To demonstrate the capability of 1 km datasets, we conducted ELM2 simulations over CONUS at the resolution of 1 km, using the newly developed 1 km land surface parameters for 2010. We used atmospheric forcing from the Global Soil Wetness Project Phase 3 (GSWP3; Kim, 2017) with a spatial resolution of 0.5º to drive ELM. The spatial homogeneity of atmospheric forcing within 0.5º grid cell guarantees that the spatial variability of ELM simulated variables (e.g., latent heat) within 0.5º grid cell is solely attributable to the heterogeneity of the 1 km land surface parameters. There are approximately 12 million effective grids over CONUS. We ran ELM for five years (2010–2014), and the last year's simulation was used for analysis. We specifically analyzed the annual mean of surface layer soil moisture (SM, $m^3/m^3$), latent heat (LH, $W/m^2$), emitted longwave radiation (ELR, $W/m^2$), and absorbed shortwave radiation (ASR, $W/m^2$).

### 3.2 Spatial scaling analysis

We conducted a spatial scaling analysis following the method described in Vergopolan (2022) on the 1 km ELM simulation data to better understand how k-scale spatial heterogeneity in the four ELM-simulated variables (mentioned in Section 3.1) induced only by spatial heterogeneity of land surface parameters changes across spatial scales. First, we performed upscaling by averaging the 1 km (=1/120°) land surface parameters and the four ELM-simulated variables to coarser spatial scales, $\lambda_{scale}$of 1/60°, 1/40°, 1/30°, 1/24°, 1/20°, and 1/10°, and calculated the spatial standard deviation ($\sigma_{scale}$) within each 0.5° × 0.5° box at each spatial scale (Table 2). Second, we quantified the changes in spatial variability at different spatial scales compared to the original 1km resolution by calculating the ratio of $\sigma_{scale}$ to $\sigma_{1\,km}$. Third, we fitted a $\log\left(\frac{\sigma_{scale}}{\sigma_{1\,\mathrm{km}}}\right) \propto \beta \times \log\left(\frac{\lambda_{scale}}{\lambda_{1\,\mathrm{km}}}\right)$ relationship, where $\beta$ is an indicator to quantify data spatial variability persistence across scales

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




(Hu et al., 1997). A more negative $\beta$ indicates a larger dependency of data spatial variability on spatial scales, resulting in a higher information loss, denoted as $\gamma = (1 - \sigma_{scale}/\sigma_{1\,\mathrm{km}}) \times 100\%$. Given the possibility that β may not demonstrate significant temporal variation (Mälicke et al., 2020), and considering that our scaling analysis is intended for demonstration purposes, our spatial scaling analysis is based on the annual mean of ELM2 simulations.

Table 2. Spatial resolution and pixel number at different spatial scales.

| $\lambda_{scale}$ | 1 | 2 | 3 | 4 | 5 | 6 | 12 |
|---|---|---|---|---|---|---|---|
| Spatial resolution | 1km (1/120°) | 2km (1/60°) | 3km (1/40°) | 4km (1/30°) | 5km (1/24°) | 6km (1/20°) | 12km (1/10°) |
| Pixel number within 0.5° × 0.5° box | 60 × 60 | 30 × 30 | 20 × 20 | 15 × 15 | 12 × 12 | 10 × 10 | 5 × 5 |

### 3.3 Attribution analysis utilizing XML methods

We conducted additional analysis to determine the primary land surface parameters that influence the spatial scaling of ELM simulations. We employed XML methods, specifically the eXtreme Gradient Boosting(XGBoost; Chen and Guestrin, 2016) machine learning algorithm and the game theoretic approach SHapley Additive exPlanations (SHAP; Lundberg and Lee, 2017; Lundberg et al., 2018, 2020). XML methods were utilized to assess the influence of land surface parameters on the spatial variability and information loss of ELM2 simulations across the CONUS. Taking spatial variability as an example, we first computed the standard deviation (σ) within each 0.5º x 0.5º grid for both 1 km resolution land surface parameters and simulations. Then, we train a machine learning model to predict the spatial variability of each simulated variable (i.e., SM, LH, ELR, ASR). We used the spatial variability (i.e., σ) and mean (μ) of the land surface parameters and μ of precipitation and temperature as predictor variables, and the simulated variable's σ as the target variable. After training the machine learning model, we used SHAP to quantify the relative importance and determine which factors were most important in driving the spatial variability of

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


the simulations. Similarly, we used this approach to identify the most critical drivers of information loss.

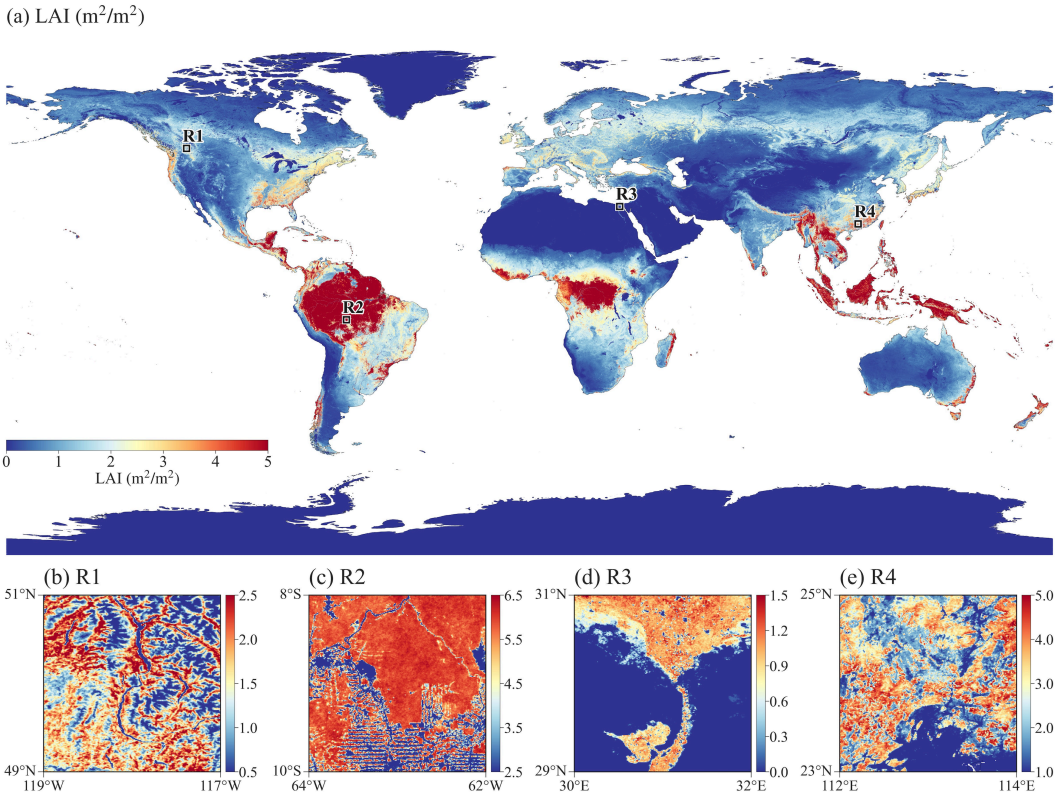
## 3. Results

### 3.1 Demonstration of the global 1km land surface parameters

LAI generally shows high values in humid and warm regions, such as tropical rainforests, southeastern US, and southern Asia, and low values over arid or cold regions, such as central Australia, southwestern US, Middle East, Central Asia, and northern Canada (Figure 1a). At high resolution, the LAI dataset 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 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).

Figure 2 demonstrates the distribution of plant functional types and other non-vegetation land units. High-resolution LULC types over multiple years can benefit studies related to LULC changes like urbanization and deforestation. Canopy height generally follows a similar spatial pattern with LAI, with high values in humid and warm regions and low values over arid or cold regions (Figure 3a). The percent clay shows high values over Southeast Asia, India, central Africa, and southeast South America, and low content over North Europe, South Africa and Alaska (Figure 3b). The topography factors follow the elevation patterns (Figures 3c and 3d), where there are large slopes and standard deviation of elevation over mountainous regions, such as the Rocky Mountains in North America, the Himalayas Mountains in Asia, and Andes Mountains in South America.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


298

299 Figure 1. The spatial pattern of LAI (annual mean in 2010) over (a) global land and (b)~(e) four

300 subregions R1~R4 within 2-degree boxes marked in (a).

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023

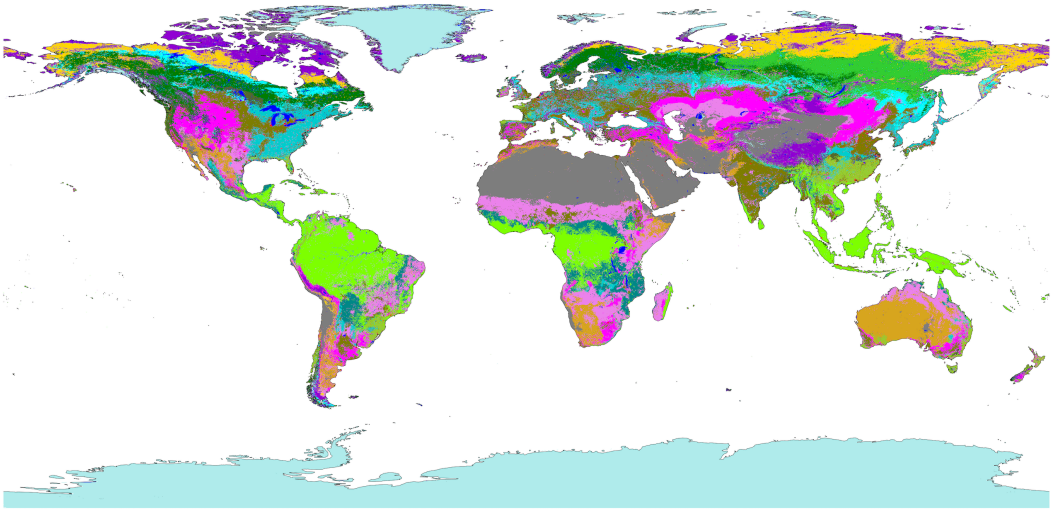




Figure 2. Global LULC distribution in year 2010.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023



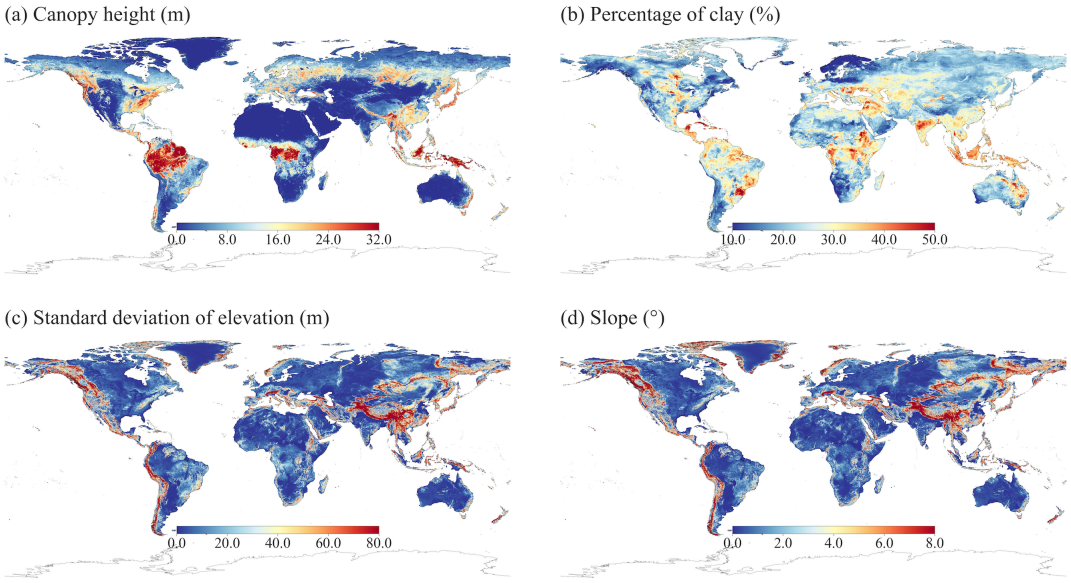




Figure 3. Demonstration of global 1km datasets (a) Canopy top height, (b) percent clay, (c) standard deviation of elevation, and (d) slope.

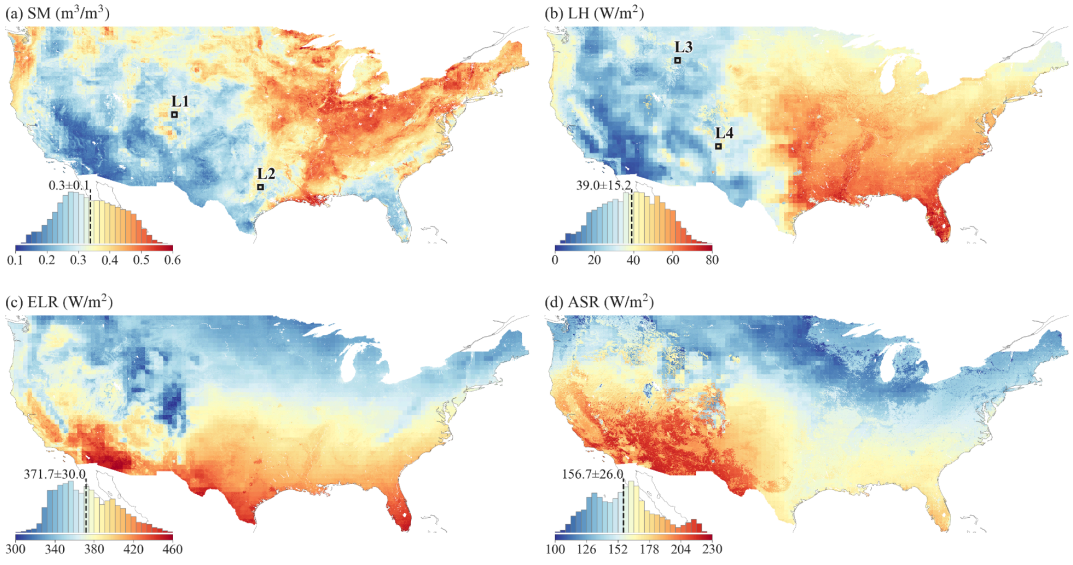
### 3.2 Demonstration 1km simulation over CONUS

ELM simulations at a 1 km resolution display significant spatial heterogeneity over CONUS (Figure 4). The values of SM, LH, ELR, and ASR across CONUS follow approximately normal distributions, with averages of 0.3 $m^3/m^3$, 39.0 $W/m^2$, 371.7 $W/m^2$, 156.7 $W/m^2$, respectively (as shown in the histogram plots in Figure 4). SM shows drier conditions over the West and Southwest and wetter conditions over the Midwest, Corn Belt, Mississippi River basin, and Northeast (Figure 4a). LH shows high values over the central and southeast, and lower values over the west and southwest (Figure 4b). The ELR generally shows higher values over regions with high surface temperature in the south (Figure 4c). The ASR shows higher values over the southwestern regions determined by incoming solar radiation and albedo (Figure 4d). Despite the high-resolution heterogeneity shown at 1 km resolution, we can still see the spatial patterns distinguished at coarse

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




319 resolution, i.e., 0.5º × 0.5º. These coarser footprints are from the GSWP3 atmospheric forcing with
320 0.5º resolution. As concluded by Li et al. (2022), atmospheric forcing is one primary heterogeneity
321 source for land surface modeling. Therefore, k-scale atmospheric forcing needs to be developed to
322 further advance k-scale offline land surface modeling.




323

324 Figure 4. The annual mean of 1 km simulations of (a)SM, (b)LH, (c) ELR, and (ASR) over
325 CONUS. The 0.5° × 0.5° boxes marked as L1, L2, L3, and L4 in (a) and (b) are selected to
326 demonstrate the spatial scaling analysis. The inserted histogram plot illustrates the distribution of
327 ELM2 simulations.

328

329 **3.3 Demonstration of spatial scaling across scales**

330 We next demonstrate the relationships between spatial variabilities and spatial scales for SM and
331 LH. Four locations (in Figures 4a and 4b) are specifically chosen to showcase varying levels of
332 spatial information loss: L1 and L3 demonstrate a relatively large loss for SM and LH, respectively,
333 while L2 and L4 represent a relatively small loss for SM and LH, respectively.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


334 At location L1 (Figure 5a), when the 1 km simulation is upscaled to coarser resolutions (i.e., larger
335 spatial scale ratios), the spatial variability of SM decreases, resulting in a negative slope of β. As
336 shown in Figure 6a, compared to the original 1 km resolution, the information loss γ reaches up to
337 54.9% at the 12 km spatial scale. The spatial pattern of SM is consistent with the spatial pattern of
338 percent clay (Figures 6a vs. 6b and 6c vs. 6d), indicating that soil texture contributes significantly
339 to the spatial variability of SM. However, SM has a more negative β than the percent clay (β = –
340 0.28 vs. –0.19 at L1, as shown in Figure 5a), suggesting that SM variability is amplified likely by
341 other processes that are also influenced by soil texture. In contrast to location L1, location L2
342 exhibits less negative β values for both SM and percent clay, suggesting that their spatial
343 variabilities exhibit less scale dependence (Figures 5a, 6c, and 6d). Both SM and percent clay at
344 location L2 approximately maintain their spatial patterns of high values in the west and low values
345 in the east across spatial scales (Figures 6c and 6d).
346 For LH, there is a more negative β value at location L3 than at location L4 (β = –0.27 at L3 vs. –
347 0.08 at L4, as shown in Figure 5b), which indicates a larger decrease of spatial variability across
348 spatial scales and lower variability persistence at location L3 than location L4 (Figure 7). The
349 spatial pattern of LH is consistent with the spatial pattern of LAI (Figures 7a vs. 7b and 7c vs. 7d)
350 at different spatial scales, suggesting that vegetation plays a significant role in the spatial
351 variability of LH. Similar to comparison between SM and soil texture, LH has a more negative β
352 than LAI (Figure 5b).

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023

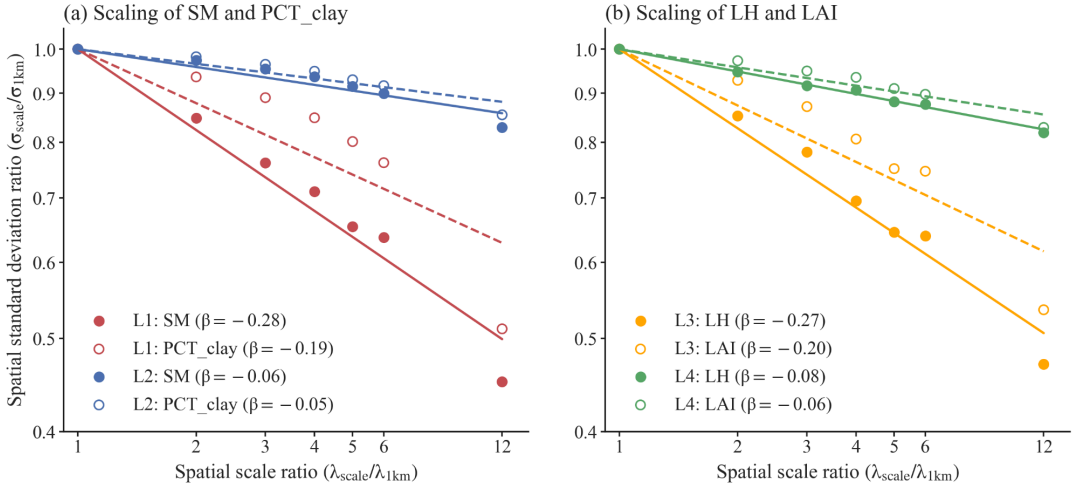




Figure 5. The scaling of spatial variabilities for (a) SM and percent clay, and (b) LH and LAI. Both the x-axis and y-axis are in logarithmic scale. The slope of the linear regression line, β, quantifies the strength of the negative relationship between spatial scale and spatial variability. A more negative β value indicates a higher spatial-scale dependency and increased information loss at coarser spatial scales. Four 0.5° × 0.5° boxes (displayed in Figure 4), namely L1 to L4, are chosen to contrast larger and smaller negative β values for SM and percent clay (L1 and L2) and for LH and LAI (L3 and L4).

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023

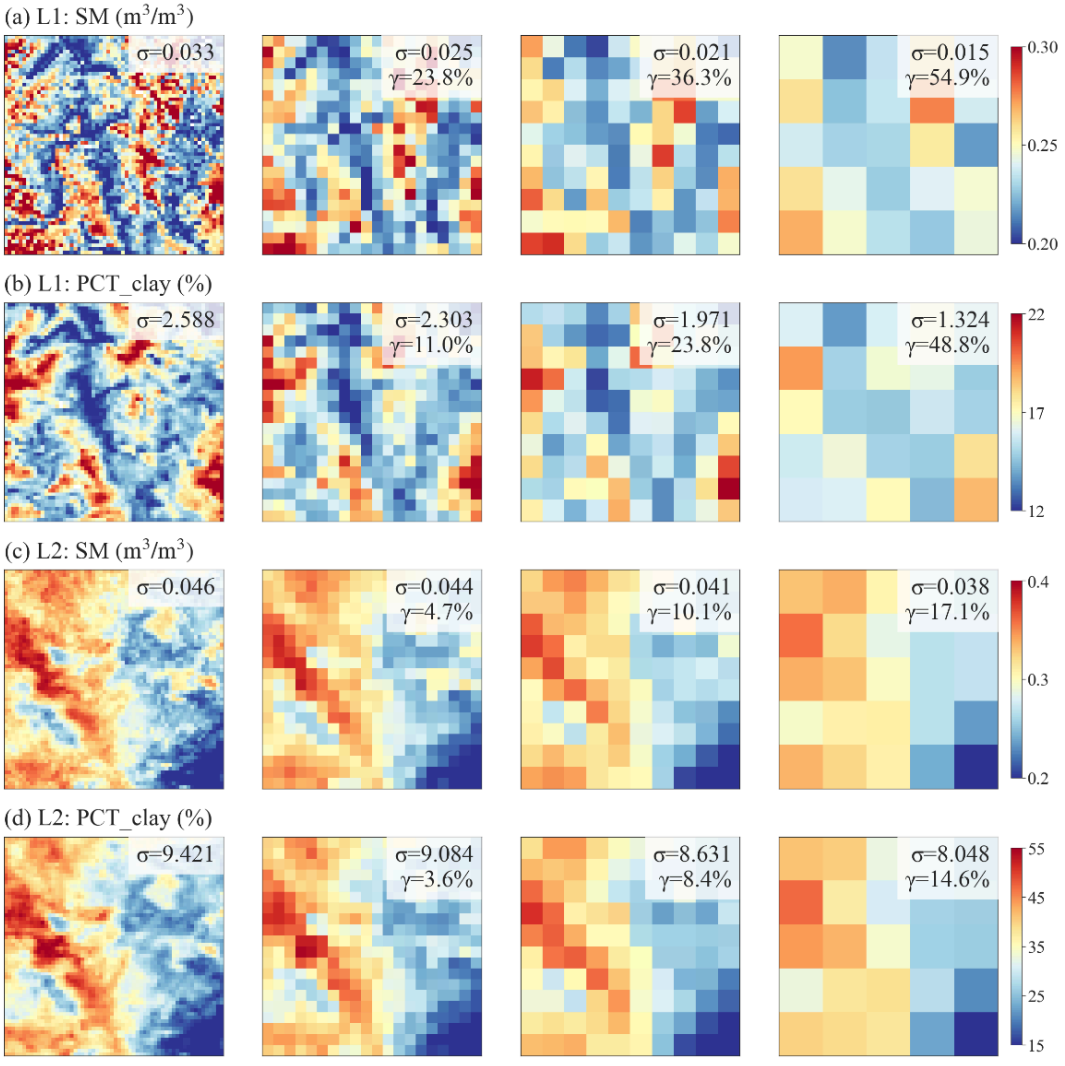




Figure 6. Comparison of SM and percent clay across spatial scales at locations L1 and L2 highlighted in Figure 5. Each subplot displays the spatial patterns of SM or percent clay within a 0.5° × 0.5° box, with the σ and γ presented in the legend.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023

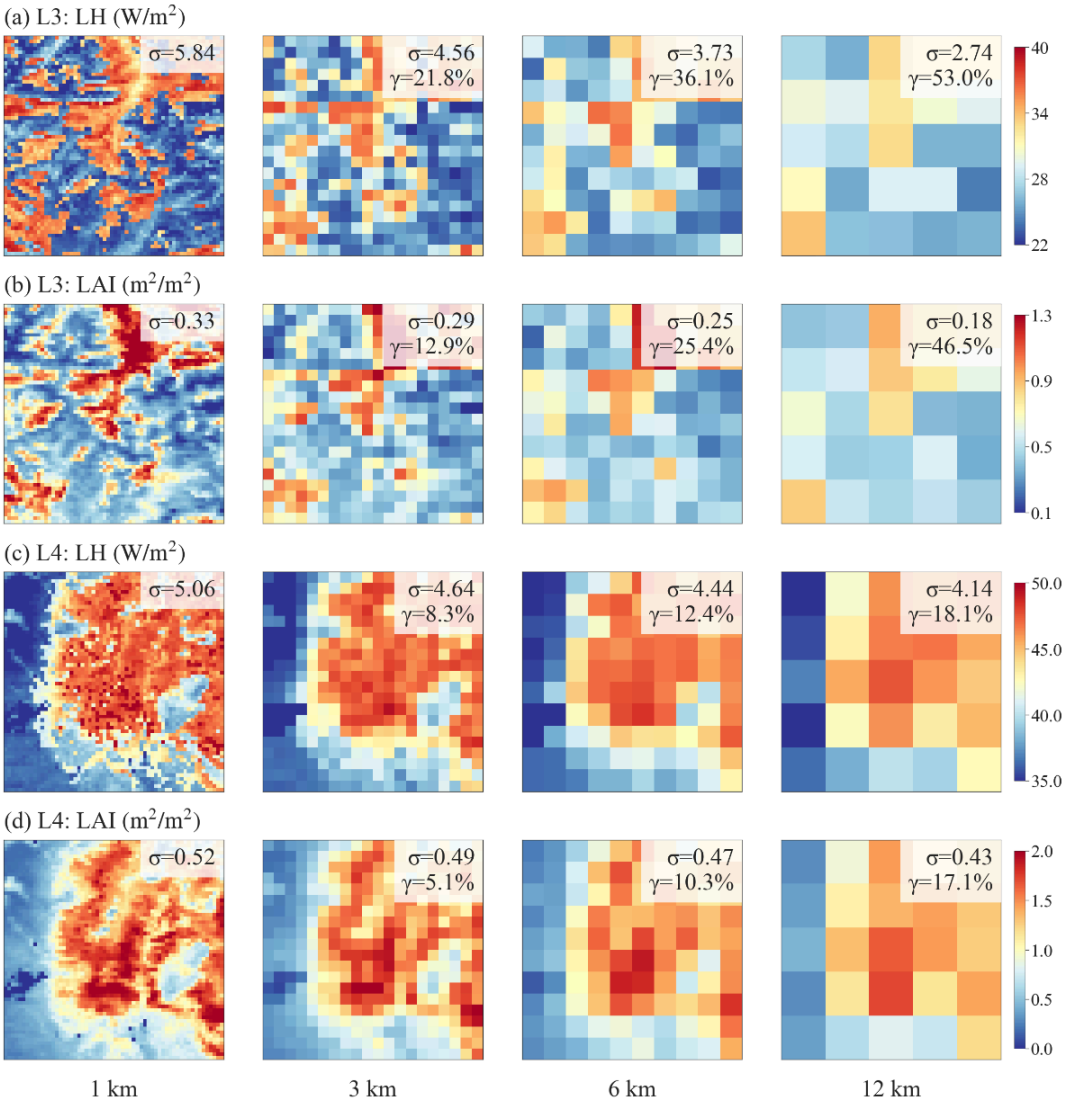




Figure 7. Similar to Figure 6, but for LH and LAI at locations L3 and L4.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


### 3.4 The spatial variability of water and energy simulations and their drivers

We quantified the spatial variability simulated at 1 km resolution using σ within each 0.5º × 0.5º box across CONUS. Four ML models were built to explore the spatial relationships between σ and its potential drivers including σ 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 $R^2$ (see Figure S1). SM shows larger spatial variability in the US Southern Coastal Plain, lower Mississippi River, Northeast, Southeast, and regions around the Great Lake (Figure 8a), which is roughly consistent with the spatial heterogeneity of the high-resolution SM simulation in Vergopolan et al. (2022). Based on the SHAP method, 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 8b). Mean precipitation and temperature reflect climate conditions (Figure S3), which are related to the water supply and water demand of soil water content. The spatial heterogeneity of soil properties, such as texture and organic matter content, affects soil hydraulic properties and generate more spatially variable soil water content. Vegetation characteristics, such as canopy height and LAI, could influence SM spatial variability through their effect on roughness length and rooting depth.

The spatial variability of LH is large in the southeastern, central, and western mountainous regions of the US (Figure 8c). Vegetation properties and climate conditions mainly drive the variability of LH (Figure 8d). 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

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


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 US, with ASR additionally showing significant spatial variability across the Northern US (Figures 8e and 8g). 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 (Figures 8f and 8h). 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.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023

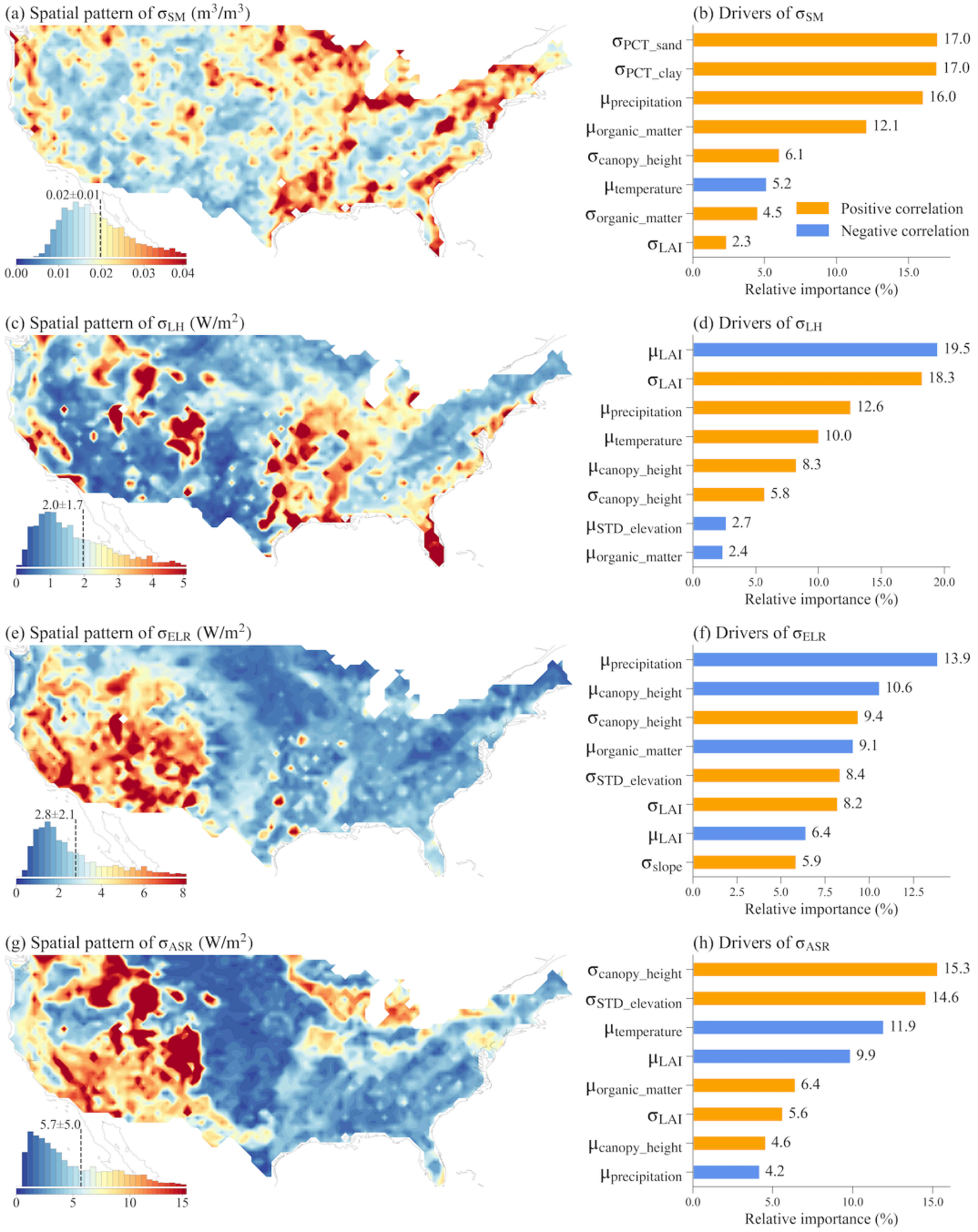




Figure 8. The spatial variability over each 0.5° × 0.5° grid cell (left plots) and the top eight most important drivers (right plots) of the spatial variability for SM, LH, ELR, and ASR. The inserted

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


401 histogram plot illustrates the probability distribution of the spatial variability across CONUS. The
402 relative importance of each variable in determining the spatial variability is calculated as the ratio
403 of the mean |SHAP value| of the variable to the sum of the mean |SHAP value| of all variables.
404 Therefore, the sum of the relative importance of all variables is 100%.
405
406 **3.5 The information loss of water and energy simulations and their drivers**
407 We also evaluated the information loss in simulations when upscaling from 1 km to 12 km
408 resolution and analyzed the drivers of their spatial patterns over CONUS. Four ML models were
409 built to explore the relationships between the γ of the simulations and its drivers including the γ of
410 the land surface parameters and the mean temperature and precipitation averaged over the 0.5º ×
411 0.5º box. These ML models performed well in predicting the simulations' γ, with small RMSE and
412 large $R^2$ (Figure S2).
413 Significant information loss ranging from 31% to 54% with maximum values exceeding 90% is
414 observed for SM, LH, ELR, and ASR simulations (Figure 9). Their spatial patterns and drivers
415 show distinct variations. $\gamma_{SM}$ is primarily driven by the information loss of percent clay and sand,
416 mean soil organic matter, and mean temperature, which affects the soil hydraulic properties and
417 soil water balance (Figures 9a and 9b). $\gamma_{LH}$ displays high values in the eastern US and low values
418 in the western US (Figure 9c). It is primarily contributed by the information loss of vegetation
419 properties such as LAI and canopy height, and mean LAI, which influences the partitioning of LH
420 and sensible heat, and the partitioning of transpiration and evaporation (Figure 9d). $\gamma_{ELR}$ exhibits
421 high values in the central and eastern US, particularly in the northeastern US, while $\gamma_{ASR}$ has high
422 values almost all over the US, especially in the eastern regions (Figures 9e and 9g). $\gamma_{ELR}$ and $\gamma_{ASR}$
423 are largely driven by vegetation properties such as LAI and canopy height, which are associated

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




424 with energy processes such as albedo (Figures 9f and 9h). Additionally, topography factors of

425 standard deviation of elevation and slope also slightly contribute to $\gamma_{ASR}$.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023

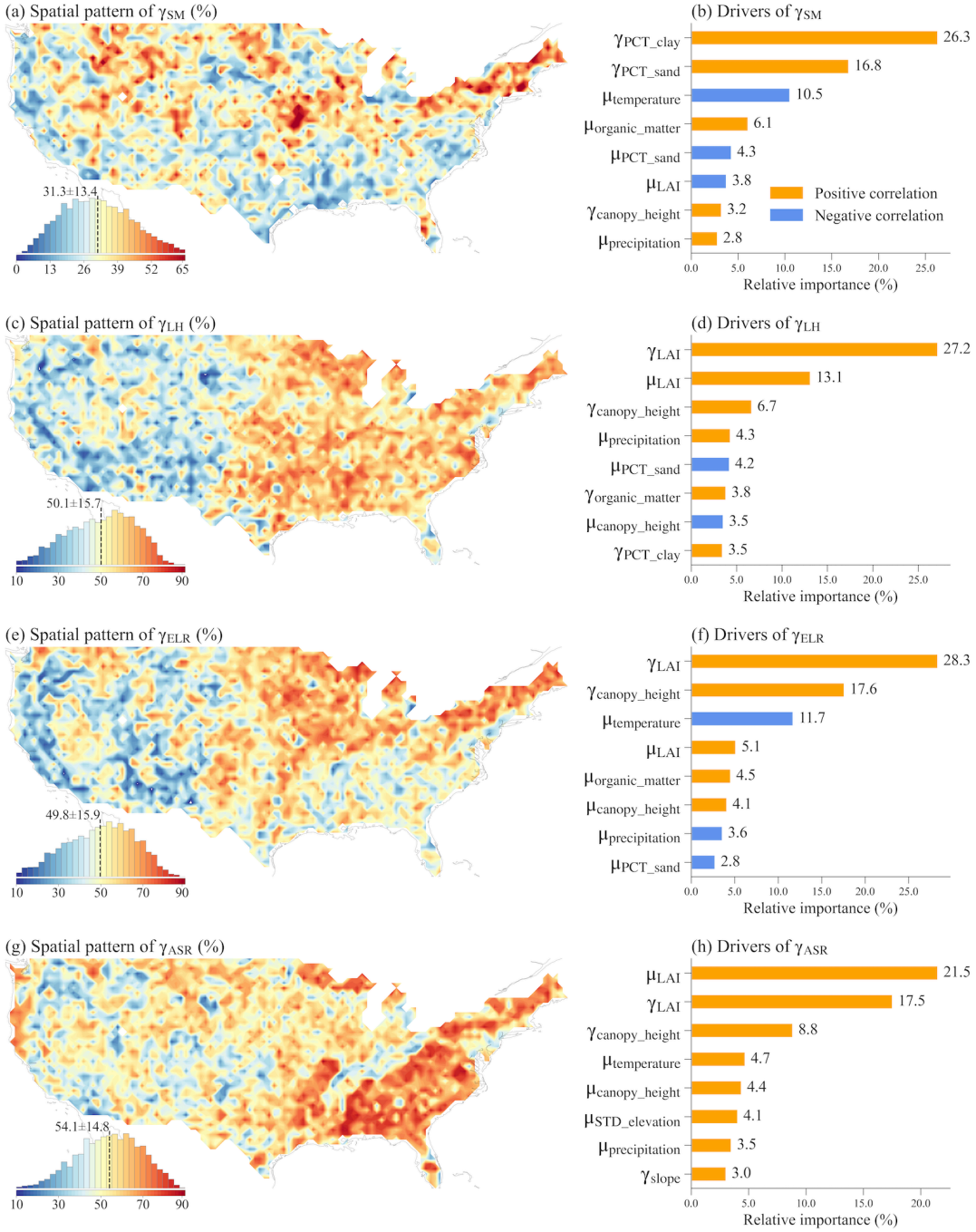




Figure 9. Same as Figure 8 but for information loss.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




## 4. Discussion

The new 1 km land surface parameter datasets developed in this study represent significant improvements over the current datasets. Compared to the two common land surface parameter datasets of CLM5 and K2012, our data are more advanced by utilizing 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. Incorporating these new 1 km land surface parameters into ESMs will advance the accurate representation and understanding of land surface processes and land–atmosphere interactions. However, certain limitations and opportunities for future development should be noted. The urban extension may differ depending on the data sources, urban definitions, and the algorithms employed, such as those derived from harmonized nighttime lights (Zhao et al., 2022), global artificial impervious area (GAIA, Li et al., 2020b; Gong et al., 2020), urban expansion (Liu et al., 2020; Kuang et al., 2021), which may differ from our urban land units based on MODIS. The urban classification into TBD, HD, and MD in J2010 is based on the global building height. Although there are building height datasets available for specific regions, such as Europe, the US, and China (Yang and Zhao, 2022; Li et al., 2020a; Frantz et al., 2021; Cao and Huang, 2021) , a globally consistent building height dataset is currently not publicly accessible, which hinders the future improvement of urban classification. The multiple-year high-resolution PFT maps developed by the European Space Agency's Climate Change Initiative could be used to further extend this dataset for a longer period (Harper et al., 2023).

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


450 The new 1 km land surface parameters can improve k-scale offline LSMs modeling by better
451 capturing spatial surface heterogeneity. As evidenced by the 1 km ELM simulation over CONUS,
452 soil properties, vegetation properties, and topographic factors contribute to the spatial
453 heterogeneities of ELM water and energy simulations. Upscaling 1 km to a coarser 12 km
454 resolution, we observe significant spatial information loss, with SM experiencing an average loss
455 of 31%, and LH, ELR, and ASR experiencing around 50% information loss on average. This
456 conclusion is in line with the results of Vergopolan et al. (2022), which showed a substantial loss
457 of spatial information in soil moisture when upscaling from 30 m to 1 km resolution, with an
458 average loss of approximately 48% and up to 80% over the CONUS region. The XML analysis
459 reveals that the spatial variability and information loss of ELM2 simulations are influenced by the
460 spatial variability and information loss of the different variables of land surface parameters, as well
461 as the mean precipitation and temperature. Our findings highlight the critical role of land surface
462 parameters in contributing to the spatial variability of water and energy in land surface simulations,
463 showcasing the value of the developed high-resolution datasets. Another implementation example
464 where our 1 km land surface parameters can be beneficial is in hillslope-scale simulations, which
465 are fundamental for organizing water, energy, and biogeochemical processes (Fan et al., 2019).
466 Krakauer et al. (2014) have highlighted the significance of between-cell groundwater flow, which
467 becomes comparable in magnitude to recharge at grid spacings smaller than 10 km. Advancements
468 have been made in ESMs to address hillslope-scale processes, including the representation of intra-
469 hillslope lateral subsurface flow within grid cells in CLM5 (Swenson et al., 2019), the development
470 of explicit lateral flow processes between grid cells (Qiu et al., 2023), and the incorporation of
471 topographic radiation effects within and between grid cells (Hao et al., 2021). Another notable
472 example is the integrated hydrology-land surface model ParFlow-CLM, which incorporates three-

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


473 dimensional groundwater flow, two-dimensional overland flow, and land surface exchange
474 processes (Maxwell, 2013). ParFlow-CLM has demonstrated remarkable reliability in reproducing
475 hydrologic processes, such as its simulations at 3 km resolution for pan-European and 1 km
476 resolution for CONUS (Naz et al., 2023; O'Neill et al., 2021). More recently, Fang et al. (2022)
477 coupled ParFlow with ELM and the Functionally Assembled Terrestrial Ecosystem Simulator
478 (FATES) to simulate carbon-hydrology interactions at hillslope scale. By incorporating our 1 km
479 datasets and leveraging these advancements, we can improve simulations of hillslope-scale
480 processes and enhance our understanding of water and energy dynamics within ESMs.
481
482 The new land surface parameters are also a timely resource for supporting the emerging need for
483 k-scale Earth system modeling. Representing the impact of spatial heterogeneity on land-
484 atmosphere interaction processes is a major challenge in Earth system modeling. Taking E3SM as
485 an example, researchers have proposed three key approaches to enhance spatial heterogeneity
486 representation to address this challenge. In line with these approaches, our newly developed 1 km
487 land surface parameters offer promising opportunities for improving land-atmosphere coupling
488 within ESMs. The first approach to enhance the representation of spatial heterogeneity is to
489 directly conduct simulations at high resolution. For instance, the Simple Cloud-Resolving E3SM
490 Atmosphere Model (SCREAM) has been used to perform global simulations at 3.25 km (Caldwell
491 et al., 2021), although the land surface parameters were based on coarser resolution datasets. By
492 utilizing the new 1 km land surface parameters, we can enhance the representation of land surface
493 heterogeneity within the ELM component of SCREAM, potentially improving modeling of land–
494 atmosphere coupling. The second and third approaches focus on improving the representation of
495 land surface heterogeneity within ESMs run at a coarse resolution while accounting for subgrid

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023


heterogeneity in two different ways. In the second approach, the Cloud Layers Unified By Binormals (CLUBB) has been implemented in E3SM Atmosphere Model (EAM) version 1 (Rasch et al., 2019; Bogenschutz et al., 2013), to better account for subgrid atmospheric heterogeneity of turbulent mixing, shallow convection, and cloud macrophysics. Recently, Huang et al. (2022) developed a novel land-atmosphere coupling scheme in EAM that enables the communication of subgrid land surface heterogeneity information to the atmosphere model with CLUBB, significantly impacting boundary layer dynamics. The new 1km datasets can provide more accurate land surface representations of the variability of individual patches and the inter-patch variability that were used in Huang et al. (2022). The third approach is the Multiple Atmosphere Multiple Land (MAML) approach used in the multiscale modeling framework (MMF) in which a cloud resolving model (CRM) is embedded within each grid cell of the atmosphere (Baker et al., 2019; Lin et al., 2023; Lee et al., 2023). In the MAML approach, each CRM column within the atmosphere grid is coupled directly with its own independent land surface. This enables a more explicit representation of the impact of spatial heterogeneity on land-atmosphere interactions within each grid and has shown notable impacts on water and energy simulations (Baker et al., 2019; Lin et al., 2023). Lee et al. (2023) highlighted the limitation of the current MAML approach, which utilizes the same land surface characteristics for each land surface model interacting with the CRM column within the same grid, which could lead to a weak representation of land-atmosphere interactions. To address this limitation, incorporating the new 1 km land surface parameters within the MAML approach can provide more detailed information about land surface heterogeneity, enabling a more accurate capture of land-atmosphere interactions.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




## 5. Data availability

The 1 km land surface parameters are publicly available at https://doi.org/10.25584/PNNLDH/1986308 (Li et al., 2023).

## 6. Conclusion

We developed 1 km global land surface parameters using the latest available datasets covering multiple years from 2001 to 2020. These parameters comprise four categories: LULC of PFTs and non-vegetative land cover, vegetation properties, soil properties, and topographic factors. The 1 km resolution ELM simulations conducted over CONUS demonstrate the valuable capabilities of the new datasets in enabling k-scale land surface modeling. Through scaling analysis of the 1 km resolution simulations within 0.5º × 0.5º boxes where spatial heterogeneity of the simulations is induced only by spatial heterogeneity of the land surface parameters, we revealed the significant impact of land surface parameters on the spatial variability of water and energy simulations. The spatial information loss of these simulations over CONUS is significant when upscaling from 1 km to a coarser 12 km resolution, with an average ranging from 31% to 54% and up to more than 90%. 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 of precipitation and temperature. Furthermore, the spatial variability of water and energy in the 1 km simulations is not dominated by the spatial heterogeneity of any land surface parameters, suggesting the usefulness of the multi-parameter high-resolution land surface parameter dataset. The availability of 1 km land surface parameters is a valuable resource that addresses the emerging needs of k-scale LSMs and ESMs modeling. By providing accurate and precise information, these

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




1 km land surface parameters will significantly enhance our understanding of the water, carbon, and energy cycles under global change.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




## Author contributions

LL, GB, and DH designed the study, processed the datasets, conducted experiments, and drafted the manuscript. LRL contributed to the conceptual design, discussion of results, and manuscript revisions.

## Acknowledgments

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. We thank Ye Liu at PNNL for providing suggestions on processing the canopy height dataset. Pacific Northwest National Laboratory is operated for the US Department of Energy by Battelle Memorial Institute under contract DE-AC05-76RL01830. DH acknowledges the support from the US DOE, Office of Science, Office of Biological and Environmental Research, Earth System Model Development program area, as part of the Climate Process Team projects.

## Financial support

This work was supported by the Regional and Global Modeling and Analysis program area of the US Department of Energy, Office of Science, Office of Biological and Environmental Research, as part of the multi-program, collaborative integrated Coastal Modeling (ICoM) project (grant no. KP1703110/75415).

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




## Competing interests

At least one of the (co-)authors is a member of the editorial board of the Earth System Science Data. The authors have no other competing interests to declare.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




## Reference

Arendt, A., Bliss , A., Bolch, T., et al.: Randolph Glacier Inventory–A Dataset of Global Glacier Outlines Version: 1.0, Global Land Ice Measurements from Space, Boulder Colorado, USA. Digital Media, 2012.

Baker, I. T., Denning, A. S., Dazlich, D. A., Harper, A. B., Branson, M. D., Randall, D. A., Phillips, M. C., Haynes, K. D., and Gallup, S. M.: Surface-Atmosphere Coupling Scale, the Fate of Water, and Ecophysiological Function in a Brazilian Forest, J Adv Model Earth Sy, 11, 2523–2546, https://doi.org/10.1029/2019ms001650, 2019.

Batjes, N.H.: ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid. Report 2006/02, available through : http://www.isric.org, 2006.

Bogenschutz, P. A., Gettelman, A., Morrison, H., Larson, V. E., Craig, C., and Schanen, D. P.: Higher-Order Turbulence Closure and Its Impact on Climate Simulations in the Community Atmosphere Model, J Climate, 26, 9655–9676, https://doi.org/10.1175/jcli-d-13-00075.1, 2013.

Bonan, G. B., Oleson, K. W., Vertenstein, M., Levis, S., Zeng, X., Dai, Y., Dickinson, R. E., and Yang, Z.-L.: The Land Surface Climatology of the Community Land Model Coupled to the NCAR Community Climate Model*, J Climate, 15, 3123–3149, https://doi.org/10.1175/1520-0442(2002)015<;3123:tlscot>2.0.co;2, 2002.

Bonan, G. B., Levis, S., Kergoat, L., & Oleson, K. W.: Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. Global Biogeochemical Cycles, 16(2), 5-1–5-23. https://doi.org/10.1029/2000gb001360, 2002

Bou-Zeid, E., Anderson, W., Katul, G. G., and Mahrt, L.: The Persistent Challenge of Surface Heterogeneity in Boundary-Layer Meteorology: A Review, Bound-lay Meteorol, 177, 227–245, https://doi.org/10.1007/s10546-020-00551-8, 2020.

Caldwell, P. M., Mametjanov, A., Tang, Q., Roekel, L. P. V., Golaz, J., Lin, W., Bader, D. C., Keen, N. D., Feng, Y., Jacob, R., Maltrud, M. E., Roberts, A. F., et al.: The DOE E3SM Coupled Model Version 1: Description and Results at High Resolution, J Adv Model Earth Sy, 11, 4095–4146, https://doi.org/10.1029/2019ms001870, 2019.

Caldwell, P. M., Terai, C. R., Hillman, B., Keen, N. D., Bogenschutz, P., Lin, W., et al.: Convection-Permitting Simulations With the E3SM Global Atmosphere Model, J Adv Model Earth Sy, 13, https://doi.org/10.1029/2021ms002544, 2021.

Cao, Y. and Huang, X.: A deep learning method for building height estimation using high-resolution multi-view imagery over urban areas: A case study of 42 Chinese cities, Remote Sens Environ, 264, 112590, https://doi.org/10.1016/j.rse.2021.112590, 2021.

Chaney, N. W., Huijgevoort, M. H. J. V., Shevliakova, E., Malyshev, S., Milly, P. C. D., Gauthier, P. P. G., and Sulman, B. N.: Harnessing big data to rethink land heterogeneity in Earth

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




604 system models, Hydrol Earth Syst Sc, 22, 3311–3330, https://doi.org/10.5194/hess-22-3311-
605 2018, 2018.

606 Change, N. C.: Think big and model small, Nat Clim Change, 12, 493–493,
607 https://doi.org/10.1038/s41558-022-01399-1, 2022.

608 Chen, J., Hagos, S., Xiao, H., Fast, J. D., and Feng, Z.: Characterization of Surface
609 Heterogeneity-Induced Convection Using Cluster Analysis, J Geophys Res Atmospheres, 125,
610 https://doi.org/10.1029/2020jd032550, 2020.

611 Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, Proc 22nd Acm Sigkdd
612 Int Conf Knowl Discov Data Min, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

613 Defries, R. S., Hansen, M. C., Townshend, J. R. G., Janetos, A. C., and Loveland, T. R.: A new
614 global 1-km dataset of percentage tree cover derived from remote sensing: GLOBAL
615 PERCENTAGE TREE COVER FROM REMOTE SENSING, Global Change Biol, 6, 247–254,
616 https://doi.org/10.1046/j.1365-2486.2000.00296.x, 2000.

617 Dozier, J.: Revisiting Topographic Horizons in the Era of Big Data and Parallel Computing, Ieee
618 Geosci Remote S, 19, 1–5, https://doi.org/10.1109/lgrs.2021.3125278, 2022.

619 Fan, Y., Clark, M., Lawrence, D. M., Swenson, S., Band, L. E., Brantley, S. L., et al.: Hillslope
620 Hydrology in Global Change Research and Earth System Modeling, Water Resour Res, 55,
621 1737–1772, https://doi.org/10.1029/2018wr023903, 2019.

622 Frantz, D., Schug, F., Okujeni, A., Navacchi, C., Wagner, W., Linden, S. van der, and Hostert,
623 P.: National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series,
624 Remote Sens Environ, 252, 112128, https://doi.org/10.1016/j.rse.2020.112128, 2021.

625 Friedl, M. A., McIver, D. K., Hodges, J. C. F., Zhang, X. Y., Muchoney, D., Strahler, A. H.,
626 Woodcock, C. E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., and Schaaf, C.:
627 Global land cover mapping from MODIS: algorithms and early results, Remote Sens Environ,
628 83, 287–302, https://doi.org/10.1016/s0034-4257(02)00078-0, 2002.

629 Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and
630 Huang, X.: MODIS Collection 5 global land cover: Algorithm refinements and characterization
631 of new datasets, Remote Sens Environ, 114, 168–182, https://doi.org/10.1016/j.rse.2009.08.016,
632 2010.

633 Friedl, M., Sulla-Menashe, D.: MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3
634 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed
635 2022-11-21 from https://doi.org/10.5067/MODIS/MCD12Q1.006, 2019.

636 Giorgi, F. and Avissar, R.: Representation of heterogeneity effects in Earth system modeling:
637 Experience from land surface modeling, Rev Geophys, 35, 413–437,
638 https://doi.org/10.1029/97rg01754, 1997.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




639 Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R.: Google Earth
640 Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202,
641 18–27. https://doi.org/10.1016/j.rse.2017.06.031, 2017.

642 Gong, P., Li, X., Wang, J., Bai, Y., Chen, B., Hu, T., Liu, X., Xu, B., Yang, J., Zhang, W., and
643 Zhou, Y.: Annual maps of global artificial impervious area (GAIA) between 1985 and 2018,
644 Remote Sens Environ, 236, 111510, https://doi.org/10.1016/j.rse.2019.111510, 2020.

645 Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth
646 Engine: Planetary-scale geospatial analysis for everyone, Remote Sens Environ, 202, 18–27,
647 https://doi.org/10.1016/j.rse.2017.06.031, 2017.

648 Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Carroll, M., Dimiceli, C., and Sohlberg, R.
649 A.: Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS
650 Vegetation Continuous Fields Algorithm, Earth Interact, 7, 1–15, https://doi.org/10.1175/1087-
651 3562(2003)007<;0001:gptcaa>2.0.co;2, 2003.

652 Hao, D., Bisht, G., Huang, M., Ma, P., Tesfa, T., Lee, W., Gu, Y., and Leung, L. R.: Impacts of
653 Sub-Grid Topographic Representations on Surface Energy Balance and Boundary Conditions in
654 the E3SM Land Model: A Case Study in Sierra Nevada, J Adv Model Earth Sy, 14,
655 https://doi.org/10.1029/2021ms002862, 2022.

656 Harper, K. L., Lamarche, C., Hartley, A., Peylin, P., Ottlé, C., Bastrikov, V., Martín, R. S.,
657 Bohnenstengel, S. I., Kirches, G., Boettcher, M., Shevchuk, R., Brockmann, C., and Defourny,
658 P.: A 29-year time series of annual 300 m resolution plant-functional-type maps for climate
659 models, Earth Syst Sci Data, 15, 1465–1499, https://doi.org/10.5194/essd-15-1465-2023, 2023.

660 Hengl, T., Jesus, J. M. de, Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A.,
661 Shangguan, W., Wright, M. N., et al.: SoilGrids250m: Global gridded soil information based on
662 machine learning, Plos One, 12, e0169748, https://doi.org/10.1371/journal.pone.0169748, 2017.

663 Hewitt, H., Fox-Kemper, B., Pearson, B., Roberts, M., and Klocke, D.: The small scales of the
664 ocean may hold the key to surprises, Nat Clim Change, 12, 496–499,
665 https://doi.org/10.1038/s41558-022-01386-6, 2022.

666 Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A.: Very high resolution
667 interpolated climate surfaces for global land areas, International Journal of Climatology, 25,
668 1965–1978, https://doi.org/10.1002/joc.1276, 2005.

669 Hu, Z., Islam, S., and Cheng, Y.: Statistical characterization of remotely sensed soil moisture
670 images, Remote Sens Environ, 61, 310–318, https://doi.org/10.1016/s0034-4257(97)89498-9,
671 1997.

672 Huang, M., Ma, P.-L., Chaney, N. W., Hao, D., Bisht, G., Fowler, M. D., Larson, V. E., and
673 Leung, L. R.: Representing surface heterogeneity in land-atmosphere coupling in E3SMv1

674 single-column model over ARM SGP during summertime, Geoscientific Model Dev Discuss,
675 2022, 1–20, https://doi.org/10.5194/gmd-2021-421, 2022.

676 Hugelius, G., Tarnocai, C., Broll, G., Canadell, J. G., Kuhry, P., and Swanson, D. K.: The
677 Northern Circumpolar Soil Carbon Database: spatially distributed datasets of soil coverage and
678 soil carbon storage in the northern permafrost regions, Earth Syst. Sci. Data, 5, 3–13,
679 https://doi.org/10.5194/essd-5-3-2013, 2013.

680 IGBP: Global Soil Data Task (IGBP-DIS, ISO-image of CD). International Geosphere-Biosphere
681 Program, PANGAEA, https://doi.org/10.1594/PANGAEA.869912, 2000.

682 Jackson, T. L., Feddema, J. J., Oleson, K. W., Bonan, G. B., and Bauer, J. T.: Parameterization
683 of Urban Characteristics for Global Climate Modeling, Ann Assoc Am Geogr, 100, 848–865,
684 https://doi.org/10.1080/00045608.2010.497328, 2010.

685 Jarvis, A., H.I. Reuter, A. Nelson, E. Guevara.: Hole-filled SRTM for the globe Version 4,
686 available from the CGIAR-CSI SRTM 90m Database: https://srtm.csi.cgiar.org, 2008.

687 Ji, P. and Yuan, X.: High-Resolution Land Surface Modeling of Hydrological Changes Over the
688 Sanjiangyuan Region in the Eastern Tibetan Plateau: 2. Impact of Climate and Land Cover
689 Change, J Adv Model Earth Sy, 10, 2829–2843, https://doi.org/10.1029/2018ms001413, 2018.

690 Ji, P., Yuan, X., Shi, C., Jiang, L., Wang, G., and Yang, K.: A Long-Term Simulation of Land
691 Surface Conditions at High Resolution over Continental China, J Hydrometeorol, 24, 285–314,
692 https://doi.org/10.1175/jhm-d-22-0135.1, 2023.

693 Ke, Y., Leung, L. R., Huang, M., Coleman, A. M., Li, H., and Wigmosta, M. S.: Development of
694 high resolution land surface parameters for the Community Land Model, Geosci Model Dev, 5,
695 1341–1362, https://doi.org/10.5194/gmd-5-1341-2012, 2012.

696 Ke, Y., Leung, L. R., Huang, M., and Li, H.: Enhancing the representation of subgrid land
697 surface characteristics in land surface models, Geosci Model Dev, 6, 1609–1622,
698 https://doi.org/10.5194/gmd-6-1609-2013, 2013.

699 Kim, H.: Global Soil Wetness Project Phase 3 Atmospheric Boundary Conditions (Experiment 1)
700 [Data set]. Data Integration and Analysis System (DIAS). https://doi.org/10.20783/DIAS.501,
701 2017.

702 Kourzeneva, E.: Global dataset for the parameterization of lakes in Numerical Weather
703 Prediction and Climate modeling. ALADIN Newsletter, No 37, July-December, 2009, F.
704 Bouttier and C. Fischer, Eds., Meteo-France, Toulouse, France, 46-53, 2009.

705 Kourzeneva, E.: External data for lake parameterization in Numerical Weather Prediction and
706 climate modeling. Boreal Environment Research, 15, 165-177, 2010.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




707 Krakauer, N. Y., Li, H., and Fan, Y.: Groundwater flow across spatial scales: importance for
708 climate modeling, Environ Res Lett, 9, 034003, https://doi.org/10.1088/1748-9326/9/3/034003,
709 2014.

710 Kuang, W., Du, G., Lu, D., Dou, Y., Li, X., Zhang, S., Chi, W., Dong, J., Chen, G., Yin, Z., Pan,
711 T., Hamdi, R., Hou, Y., Chen, C., Li, H., and Miao, C.: Global observation of urban expansion
712 and land-cover dynamics using satellite big-data, Sci Bull, 66, 297–300,
713 https://doi.org/10.1016/j.scib.2020.10.022, 2021.

714 Lang, N., Jetz, W., Schindler, K., and Wegner, J. D.: A high-resolution canopy height model of
715 the Earth, Arxiv, https://doi.org/10.48550/arxiv.2204.08322, 2022.

716 Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al.:
717 The Community Land Model Version 5: Description of New Features, Benchmarking, and
718 Impact of Forcing Uncertainty, J Adv Model Earth Sy, 11, 4245–4287,
719 https://doi.org/10.1029/2018ms001583, 2019.

720 Lawrence, D., Fisher, R., Koven, C., Oleson, K., Swenson, S., et al. (2018). Technical
721 description of version 5.0 of the Community Land Model (CLM). National Center for
722 Atmospheric Research, University Corporation for Atmospheric Research, Boulder, CO.
723 https://escomp.github.io/ctsm-docs/versions/release-clm5.0/html/tech_note/index.html

724 Lee, J., Hannah, W. M., and Bader, D. C.: Representation of atmosphere induced heterogeneity
725 in land – atmosphere interactions in E3SM-MMFv2, Geoscientific Model Dev Discuss, 2023, 1–
726 21, https://doi.org/10.5194/gmd-2023-55, 2023.

727 Leng, G., Huang, M., Tang, Q., Sacks, W. J., Lei, H., and Leung, L. R.: Modeling the effects of
728 irrigation on land surface fluxes and states over the conterminous United States: Sensitivity to
729 input data and model parameters, J Geophys Res Atmospheres, 118, 9789–9803,
730 https://doi.org/10.1002/jgrd.50792, 2013.

731 Leung, L. R., Bader, D. C., Taylor, M. A., and McCoy, R. B.: An Introduction to the E3SM
732 Special Collection: Goals, Science Drivers, Development, and Analysis, J Adv Model Earth Sy,
733 12, https://doi.org/10.1029/2019ms001821, 2020.

734 Li, L., Bisht, G., and Leung, L. R.: Spatial heterogeneity effects on land surface modeling of
735 water and energy partitioning, Geosci Model Dev, 15, 5489–5510, https://doi.org/10.5194/gmd-
736 15-5489-2022, 2022.

737 Li, X., Zhou, Y., Gong, P., Seto, K. C., and Clinton, N.: Developing a method to estimate
738 building height from Sentinel-1 data, Remote Sens Environ, 240, 111705,
739 https://doi.org/10.1016/j.rse.2020.111705, 2020a.

740 Li, X., Gong, P., Zhou, Y., Wang, J., Bai, Y., Chen, B., Hu, T., Xiao, Y., et al.: Mapping global
741 urban boundaries from the global artificial impervious area (GAIA) data, Environ Res Lett, 15,
742 094044, https://doi.org/10.1088/1748-9326/ab9be3, 2020b.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




743 Li, L., Bisht, G., Hao, D., Leung, L.R.: Global 1km Land Surface Parameters for Kilometer
744 Scale Earth System Modeling, Pacific Northwest National Laboratory DataHub [data set],
745 https://doi.org/10.25584/PNNLDH/1986308, 2023.

746 Lin, G., Leung, L. R., Lee, J., Harrop, B. E., Baker, I. T., Branson, M. D., Denning, A. S., Jones,
747 C. R., Ovchinnikov, M., Randall, D. A., and Yang, Z.: Modeling Land-Atmosphere Coupling at
748 Cloud-Resolving Scale Within the Multiple Atmosphere Multiple Land (MAML) Framework in
749 SP-E3SM, J Adv Model Earth Sy, 15, https://doi.org/10.1029/2022ms003101, 2023.

750 Liu, S., Shao, Y., Kunoth, A., and Simmer, C.: Impact of surface-heterogeneity on atmosphere
751 and land-surface interactions, Environ Modell Softw, 88, 35–47,
752 https://doi.org/10.1016/j.envsoft.2016.11.006, 2017.

753 Liu, X., Huang, Y., Xu, X., Li, X., Li, X., Ciais, P., Lin, P., et al.: High-spatiotemporal-
754 resolution mapping of global urban change from 1985 to 2015, Nat Sustain, 3, 564–570,
755 https://doi.org/10.1038/s41893-020-0521-x, 2020.

756 Lundberg, S. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, Arxiv, 2017.

757 Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., Liston, D. E.,
758 Low, D. K.-W., Newman, S.-F., Kim, J., and Lee, S.-I.: Explainable machine-learning
759 predictions for the prevention of hypoxaemia during surgery, Nat Biomed Eng, 2, 749–760,
760 https://doi.org/10.1038/s41551-018-0304-0, 2018.

761 Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R.,
762 Himmelfarb, J., Bansal, N., and Lee, S.-I.: From local explanations to global understanding with
763 explainable AI for trees, Nat Mach Intell, 2, 56–67, https://doi.org/10.1038/s42256-019-0138-9,
764 2020.

765 Mälicke, M., Hassler, S. K., Blume, T., Weiler, M., and Zehe, E.: Soil moisture: variable in
766 space but redundant in time, Hydrol Earth Syst Sc, 24, 2633–2653, https://doi.org/10.5194/hess-
767 24-2633-2020, 2020.

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

771 Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song,
772 X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R.,
773 and Running, S. W.: Global products of vegetation leaf area and fraction absorbed PAR from
774 year one of MODIS data, Remote Sens Environ, 83, 214–231, https://doi.org/10.1016/s0034-
775 4257(02)00074-3, 2002.

776 Myneni, R., Knyazikhin, Y., Park, T. (2021). MODIS/Terra+Aqua Leaf Area Index/FPAR 4-Day
777 L4 Global 500m SIN Grid V061 [Data set]. NASA EOSDIS Land Processes DAAC. Accessed
778 2022-11-21 from https://doi.org/10.5067/MODIS/MCD15A3H.061

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




779 Naz, B. S., Sharples, W., Ma, Y., Goergen, K., and Kollet, S.: Continental-scale evaluation of a
780 fully distributed coupled land surface and groundwater model, ParFlow-CLM (v3.6.0), over
781 Europe, Geosci Model Dev, 16, 1617–1639, https://doi.org/10.5194/gmd-16-1617-2023, 2023.

782 Nitta, T., Arakawa, T., Hatono, M., Takeshima, A., & Yoshimura, K: Development of integrated
783 land simulator, Prog Earth Planet Sci, 7(1), 1-14, https://doi.org/10.1186/s40645-020-00383-7,
784 2020.

785 O'Neill, M. M. F., Tijerina, D. T., Condon, L. E., and Maxwell, R. M.: Assessment of the
786 ParFlow–CLM CONUS 1.0 integrated hydrologic model: evaluation of hyper-resolution water
787 balance components across the contiguous United States, Geosci Model Dev, 14, 7223–7254,
788 https://doi.org/10.5194/gmd-14-7223-2021, 2021.

789 Poggio, L., Sousa, L. M. de, Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and
790 Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial
791 uncertainty, Soil, 7, 217–240, https://doi.org/10.5194/soil-7-217-2021, 2021.

792 Qiu, H., Bisht, G., Li, L., Hao, D., and Xu, D.: Development of Inter-Grid Cell Lateral
793 Unsaturated and Saturated Flow Model in the E3SM Land Model (v2.0), Egusphere, 2023, 1–31,
794 https://doi.org/10.5194/egusphere-2023-375, 2023.

795 Rabus, B., Eineder, M., Roth, A., & Bamler, R.: The shuttle radar topography mission—A new
796 class of digital elevation models acquired by spaceborne radar. ISPRS Journal of
797 Photogrammetry and Remote Sensing, 57(4), 241–262. https://doi.org/10.1016/s0924-
798 2716(02)00124-7, 2003.

799 Ramankutty, N. and Foley, J. A.: Estimating historical changes in global land cover: Croplands
800 from 1700 to 1992, Global Biogeochem Cy, 13, 997–1027,
801 https://doi.org/10.1029/1999gb900046, 1999.

802 Rasch, P. J., Xie, S., Ma, P. -L., Lin, W., Wang, H., Tang, Q., Burrows, S. M., Caldwell, P., et
803 al.: An Overview of the Atmospheric Component of the Energy Exascale Earth System Model, J
804 Adv Model Earth Sy, 11, 2377–2411, https://doi.org/10.1029/2019ms001629, 2019.

805 Rastner, P., Bolch, T., Mölg, N., Machguth, H., Le Bris, R., and Paul, F.: The first complete
806 inventory of the local glaciers and ice caps on Greenland, The Cryosphere, 6, 1483–1495,
807 https://doi.org/10.5194/tc-6-1483-2012, 2012.

808 Rouf, T., Maggioni, V., Mei, Y., and Houser, P.: Towards hyper-resolution land-surface
809 modeling of surface and root zone soil moisture, J Hydrol, 594, 125945,
810 https://doi.org/10.1016/j.jhydrol.2020.125945, 2021.

811 Simard, M., Pinto, N., Fisher, J. B., and Baccini, A.: Mapping forest canopy height globally with
812 spaceborne lidar, Journal of Geophysical Research: Biogeosciences, 116, G04021,
813 https://doi.org/10.1029/2011jg001708, 2011.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




814 Singh, R. S., Reager, J. T., Miller, N. L., and Famiglietti, J. S.: Toward hyper-resolution land-
815 surface modeling: The effects of fine-scale topography and soil texture on CLM4.0 simulations
816 over the Southwestern U.S., Water Resour Res, 51, 2648–2667,
817 https://doi.org/10.1002/2014wr015686, 2015.

818 Slingo, J., Bates, P., Bauer, P., Belcher, S., Palmer, T., Stephens, G., Stevens, B., Stocker, T.,
819 and Teutsch, G.: Ambitious partnership needed for reliable climate prediction, Nat Clim Change,
820 12, 499–503, https://doi.org/10.1038/s41558-022-01384-8, 2022.

821 Still, C. J., Berry, J. A., Collatz, G. J., and DeFries, R. S.: Global distribution of C3 and C4
822 vegetation: Carbon cycle implications, Global Biogeochem Cy, 17, 6-1-6–14,
823 https://doi.org/10.1029/2001gb001807, 2003.

824 Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P., and Friedl, M. A.: Hierarchical mapping of
825 annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product, Remote
826 Sens Environ, 222, 183–194, https://doi.org/10.1016/j.rse.2018.12.013, 2019.

827 Swenson, S. C., Clark, M., Fan, Y., Lawrence, D. M., and Perket, J.: Representing Intrahillslope
828 Lateral Subsurface Flow in the Community Land Model, J Adv Model Earth Sy, 11, 4044–4065,
829 https://doi.org/10.1029/2019ms001833, 2019.

830 Verdin, K. L. and Greenlee, S. K.: Development of continental scale digital elevation models and
831 extraction of hydrographic features, paper presented at the Third International Workshop on
832 Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, 21–26 January, Natl.
833 Cent. for Geogr. Inf. and Anal., Santa Barbara, Calif, 1996.

834 Vergopolan, N., Chaney, N. W., Beck, H. E., Pan, M., Sheffield, J., Chan, S., and Wood, E. F.:
835 Combining hyper-resolution land surface modeling with SMAP brightness temperatures to
836 obtain 30-m soil moisture estimates, Remote Sens Environ, 242, 111740,
837 https://doi.org/10.1016/j.rse.2020.111740, 2020.

838 Vergopolan, N., Chaney, N. W., Pan, M., Sheffield, J., Beck, H. E., Ferguson, C. R., Torres-
839 Rojas, L., Sadri, S., and Wood, E. F.: SMAP-HydroBlocks, a 30-m satellite-based soil moisture
840 dataset for the conterminous US, Sci Data, 8, 264, https://doi.org/10.1038/s41597-021-01050-2,
841 2021.

842 Vergopolan, N., Sheffield, J., Chaney, N. W., Pan, M., Beck, H. E., Ferguson, C. R., Torres-
843 Rojas, L., Eigenbrod, F., Crow, W., and Wood, E. F.: High-Resolution Soil Moisture Data
844 Reveal Complex Multi-Scale Spatial Variability Across the United States, Geophys Res Lett, 49,
845 https://doi.org/10.1029/2022gl098586, 2022.

846 Vrese, P. de, Schulz, J.-P., and Hagemann, S.: On the Representation of Heterogeneity in Land-
847 Surface–Atmosphere Coupling, Bound-lay Meteorol, 160, 157–183,
848 https://doi.org/10.1007/s10546-016-0133-1, 2016.

https://doi.org/10.5194/essd-2023-242
Preprint. Discussion started: 19 July 2023




849 Wood, E. F., Roundy, J. K., Troy, T. J., Beek, L. P. H. van, Bierkens, M. F. P., et al.:
850 Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's
851 terrestrial water, Water Resour Res, 47, https://doi.org/10.1029/2010wr010090, 2011.

852 Xia, Y., Mocko, D., Huang, M., Li, B., Rodell, M., Mitchell, K. E., Cai, X., and Ek, M. B.:
853 Comparison and Assessment of Three Advanced Land Surface Models in Simulating Terrestrial
854 Water Storage Components over the United States, J Hydrometeorol, 18, 625–649,
855 https://doi.org/10.1175/jhm-d-16-0112.1, 2017.

856 Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., and Pavelsky, T. M.: MERIT
857 Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset,
858 Water Resour Res, 55, 5053–5073, https://doi.org/10.1029/2019wr024873, 2019.

859 Yang, C. and Zhao, S.: A building height dataset across China in 2017 estimated by the spatially-
860 informed approach, Sci Data, 9, 76, https://doi.org/10.1038/s41597-022-01192-x, 2022.

861 Yuan, H., Dai, Y., Xiao, Z., Ji, D., and Shangguan, W.: Reprocessing the MODIS Leaf Area
862 Index products for land surface and climate modelling, Remote Sens Environ, 115, 1171–1187,
863 https://doi.org/10.1016/j.rse.2011.01.001, 2011.

864 Yuan, X., Ji, P., Wang, L., Liang, X., Yang, K., Ye, A., Su, Z., and Wen, J.: High-Resolution
865 Land Surface Modeling of Hydrological Changes Over the Sanjiangyuan Region in the Eastern
866 Tibetan Plateau: 1. Model Development and Evaluation, J Adv Model Earth Sy, 10, 2806–2828,
867 https://doi.org/10.1029/2018ms001412, 2018.

868 Zeng, X., Shaikh, M., Dai, Y., Dickinson, R. E., and Myneni, R.: Coupling of the Common Land
869 Model to the NCAR Community Climate Model, J Climate, 15, 1832–1854,
870 https://doi.org/10.1175/1520-0442(2002)015<;1832:cotclm>2.0.co;2, 2002.

871 Zhao, M., Cheng, C., Zhou, Y., Li, X., Shen, S., and Song, C.: A global dataset of annual urban
872 extents (1992–2020) from harmonized nighttime lights, Earth Syst Sci Data, 14, 517–534,
873 https://doi.org/10.5194/essd-14-517-2022, 2022.

874 Zhou, Y., Li, D., and Li, X.: The Effects of Surface Heterogeneity Scale on the Flux Imbalance
875 under Free Convection, J Geophys Res Atmospheres, 124, 8424–8448,
876 https://doi.org/10.1029/2018jd029550, 2019.